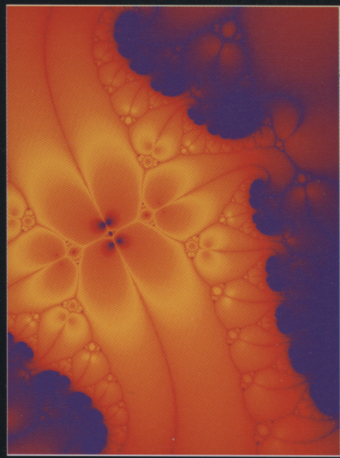
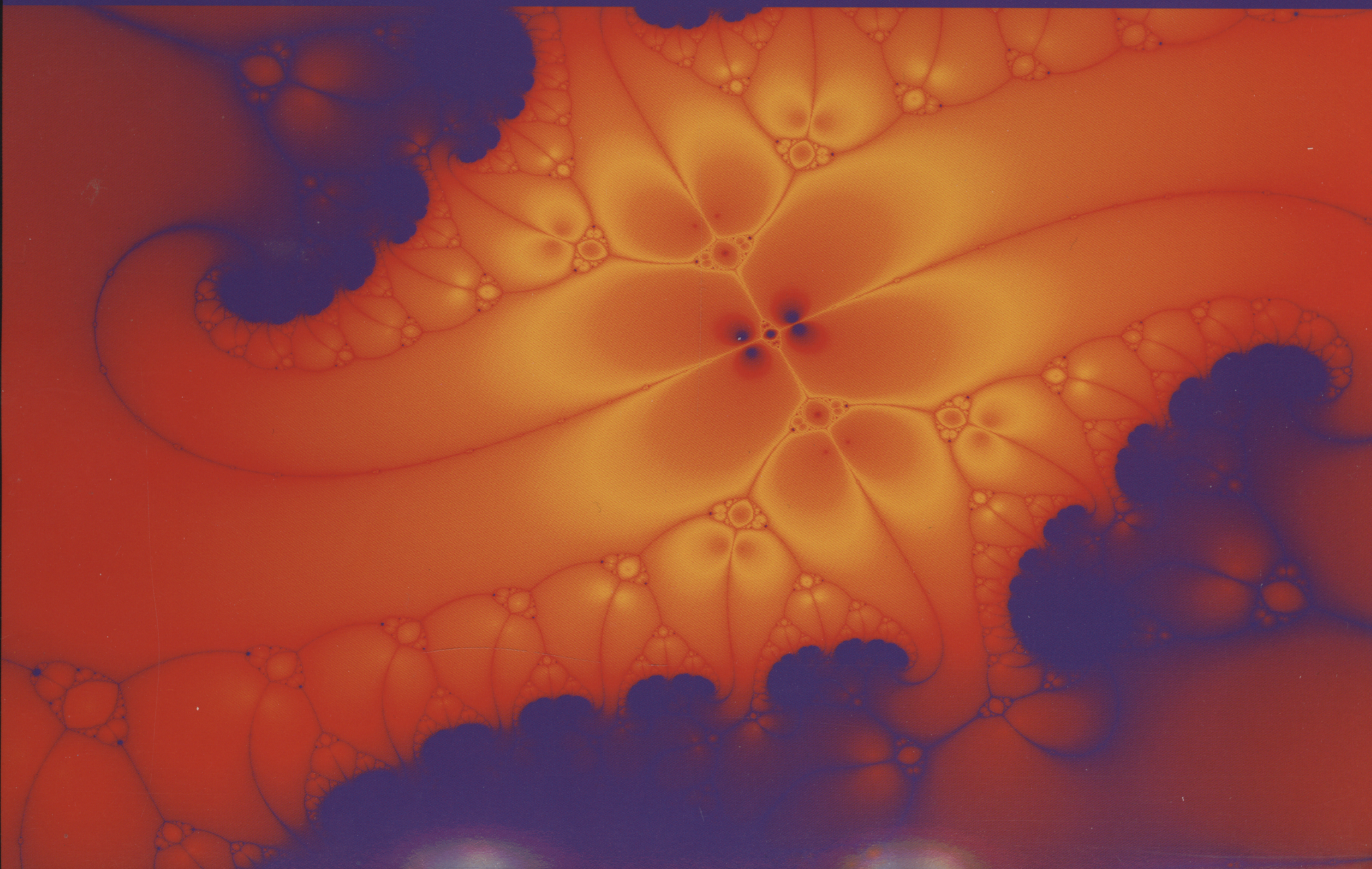
The Open University

M338 Topology



A2

Unit A2 Metric spaces



The Open University

M338 Topology

A2

Metric spaces

This publication forms part of an Open University course. Details of this and other Open University courses can be obtained from the Student Registration and Enquiry Service, The Open University, PO Box 197, Milton Keynes, MK7 6BJ, United Kingdom: tel. +44 (0)870 333 4340, e-mail general-enquiries@open.ac.uk

Alternatively, you may visit the Open University website at http://www.open.ac.uk where you can learn more about the wide range of courses and packs offered at all levels by The Open University.

To purchase a selection of Open University course materials, visit the webshop at www.ouw.co.uk, or contact Open University Worldwide, Michael Young Building, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom, for a brochure: tel. +44 (0)1908 858785, fax +44 (0)1908 858787, e-mail ouwenq@open.ac.uk

The Open University, Walton Hall, Milton Keynes, MK7 6AA.

First published 2006.

Edited, designed and typeset by The Open University, using the Open University TEX System.

Printed and bound in the United Kingdom by The Charlesworth Group, Wakefield.

ISBN 0 7492 4126 8

1.1

Contents

Introduction

This unit is the midway point in our journey from understanding continuity on the real line to understanding it in a very general setting.

In *Unit A1, Distance and Continuity*, we saw how the ε–δ definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ can be extended to a definition of continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. A key role is played in this definition by the Euclidean *distance* between pairs of points in these spaces. In this unit, we investigate how to extend this idea to a notion of continuity between arbitrary spaces in which a satisfactory notion of distance is available.

In *Unit A1*, we found that the Euclidean distance function satisfies three properties that we would clearly require any distance function to satisfy. The key idea in this unit is to raise these properties to the status of axioms and to say that any function that satisfies them is a well-defined distance function, known as a *metric*. We shall see how the idea of a metric allows our definition of continuity in Euclidean spaces to be extended naturally to other spaces.

We shall also see how the concept of a metric leads to different ways of measuring 'distance' on the same space. Sometimes two different specifications of 'distance' give rise to the *same* continuous functions, while sometimes they produce *different* ones. The fundamental concept of an *open set* enables us to recognize continuous functions and distinguish between different notions of continuity on the same space. This generalized concept of continuity is the culmination of the unit.

Study guide

In Section 1, *Introducing metric spaces*, we recall the three properties of the Euclidean distance function and show how to use them to define the notions of a *metric* and a *metric space*. We also give some examples of metric spaces, and show how the definition of continuity extends to this more general setting. The concepts described are central to the theory of metric spaces and you may find it helpful to reread this section after your study of Section 2.

Section 2, *Examples of metrics*, contains some more examples of metric spaces and functions defined upon them. In this section, we introduce techniques for showing that a given distance function is a metric, and we demonstrate how to prove that a given function is continuous. You should take time to understand these techniques.

Section 3, *New metrics from old*, describes methods for using existing metrics to define new metrics on a space. You will see that different metrics can give rise to the same continuous functions; this is the first indication that there may be a deeper notion underlying the definition of continuity. You should ensure that you understand how to transfer a metric and take the product of two metrics.

Section 4, *Open sets*, investigates why different metrics may give rise to the same continuous functions. The key role is played by a particular class of sets, known as *open sets.* We find that the collection of open sets in a metric space always has certain properties; these properties form the starting point for the next unit, *Topological Spaces.* The idea of open sets is fundamental to the definition of topological spaces, and it is important that you study this section carefully. You may find it helpful to re-view the course DVD after finishing this section.

There is no software associated with this unit.

1 Introducing metric spaces

After working through this section, you should be able to:

- explain the geometric meaning of each of the *metric space* properties (M1)–(M3);
- state the definition of *continuity* of a function between two metric spaces;
- distinguish between *open* and *closed balls* in a metric space.

This section introduces the idea of a metric space, and shows how this concept allows us to generalize the notion of continuity. Subsection 1.1 shows how properties of the Euclidean distance function on $\mathbb{R}^n$ can be generalized to define a metric on any set, and thus a metric space. Subsection 1.2 extends the definition of continuity from Euclidean spaces to metric spaces, and introduces the notions of open and closed balls.

You should not expect to have a firm grasp of the idea of a metric space by the end of this section: this will come as you see more examples of metric spaces in Sections 2 and 3.

1.1 The definition of a metric space

In *Unit A1*, we gave a definition of continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ which depends on the Euclidean distance functions, $d^{(n)}$ for $\mathbb{R}^n$ and $d^{(m)}$ for $\mathbb{R}^m$. To define a notion of continuity for functions between arbitrary sets, we must first find an effective notion of distance appropriate to any set X. To do this, an obvious strategy is to try to generalize the notion of Euclidean distance.

Recall that, if $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$, then $d^{(k)}(\mathbf{a}, \mathbf{b})$ is given by $\sqrt{\sum_{i=1}^{k}(b_i - a_i)^2}$.

In the previous unit, we showed that the Euclidean distance function has three properties.

The *Euclidean distance function* $d^{(n)}: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ has the following properties.

For all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$:

(M1) $d^{(n)}(\mathbf{a}, \mathbf{b}) \geq 0$, with equality holding if and only if $\mathbf{a} = \mathbf{b}$;

(M2) $d^{(n)}(\mathbf{a}, \mathbf{b}) = d^{(n)}(\mathbf{b}, \mathbf{a})$;

(M3) $d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c})$ (the *Triangle Inequality*).

This is Theorem 5.1 of *Unit A1*.

(M1)–(M3) do not use any special properties of the set $\mathbb{R}^n$, so it seems reasonable to say that, for any set X, a function $d: X \times X \to \mathbb{R}$ is a *distance function* if it satisfies them. It turns out that these are indeed the appropriate properties of distance on which to base our definition. We call such a distance function a *metric*, and we call a set with such a function defined on it a *metric space*.

The word *metric* comes from the Greek word *μετρόν* (metron), meaning distance.

> **Definition**
>
> Let X be a set. A **metric** on X is a function $d: X \times X \to \mathbb{R}$ satisfying the following three conditions.
>
> For all $a, b, c \in X$:
>
> (M1) $d(a,b) \geq 0$, with equality holding if and only if $a = b$;
>
> (M2) $d(a,b) = d(b,a)$;
>
> (M3) $d(a,c) \leq d(a,b) + d(b,c)$ (the **Triangle Inequality**).
>
> The set X, together with a metric d on X, is a **metric space**, and is denoted by (X, d).

This definition was proposed by Maurice Fréchet in his doctoral thesis 'Sur quelques points du calcul fonctionnel', published in the Italian journal *Rendiconti Circolo Mat. Palermo* in 1906.

Remarks

(i) Mathematicians usually refer to the members of the set X as **points**, to emphasize the analogy with the points of a line, a plane or three-dimensional space. Similarly mathematicians usually refer to $d(a,b)$ as the **distance** between a and b. The theory of metric spaces is the study of those properties of sets of points that depend only on distance.

A 'point' may be nothing like a dot in the plane: it could, for example, be a function, and X could be a set of functions.

(ii) Condition (M1) says that distance is a non-negative quantity, and that the only point of a metric space that is at zero distance from a given point is that point itself.

(iii) Condition (M2) says that the distance from a point a to a point b is precisely the same as the distance from b to a. (The metric is *symmetric*.)

(iv) Condition (M3) tells us that $d(a,c)$ gives the 'shortest' distance between a and c. For if we go directly from a to c, that gives a distance of $d(a,c)$. However, if we make a detour to b along the way, we must go a distance $d(a,b)$ to get to b, and then an additional distance $d(b,c)$ to get from b to c. The total distance travelled must then be at least as great as the 'direct' distance $d(a,c)$.

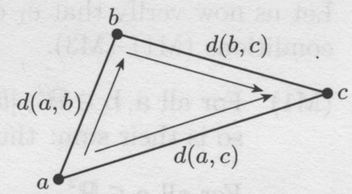


Figure 1.1

(v) As our general definition is modelled on the corresponding properties of the Euclidean distance function $d^{(n)}$, it follows that, for each $n \in \mathbb{N}$, $(\mathbb{R}^n, d^{(n)})$ is a metric space. It is known as **Euclidean *n*-space**. Furthermore, in the context of metric spaces, the Euclidean distance function $d^{(n)}$ is often referred to as the **Euclidean metric**.

In *Unit A1*, we saw that the Euclidean metric on $\mathbb{R}^n$ satisfies the Reverse Triangle Inequality, and that the proof makes use only of properties (M1)–(M3) of the Euclidean metric. So it is no surprise to learn that there is a Reverse Triangle Inequality for any metric space.

> ***Theorem 1.1 Reverse Triangle Inequality***
>
> Let (X, d) be a metric space. For all $a, b, c \in X$,
>
> (M3a) $d(b,c) \geq |d(a,c) - d(a,b)|$.

The proof is essentially the same as that given in *Unit A1*, Theorem 5.2, and so we omit it here.

The taxicab metric

We have already observed that the usual definition of distance on the plane defines a metric. Are there any others?

The usual definition is given by

$$d^{(2)}(\mathbf{a},\mathbf{b}) = \sqrt{(b_1-a_1)^2+(b_2-a_2)^2} \quad \text{for } \mathbf{a},\mathbf{b}\in\mathbb{R}^2.$$

There are other ways of defining distance on $\mathbb{R}^2$ that are mathematically simpler. For example, consider the function $e_1\colon \mathbb{R}^2\times\mathbb{R}^2\to\mathbb{R}$ defined by

$$e_1(\mathbf{a},\mathbf{b}) = |b_1-a_1|+|b_2-a_2| \quad \text{for } \mathbf{a},\mathbf{b}\in\mathbb{R}^2.$$

The reason for the subscript 1 in the notation e_1 will become clear in Subsection 3.2.

As we shall shortly see, this *does* define an alternative metric on the plane. First let us try to understand how this distance function behaves.

Problem 1.1

Find:

(a) $e_1((0,0),(1,0))$; (b) $e_1((0,0),(0,1))$; (c) $e_1((0,1),(1,0))$.

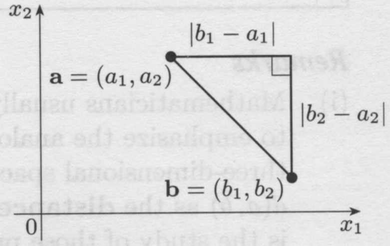


Figure 1.2

You may have been able to deduce from your work on Problem 1.1 that one way to understand $e_1(\mathbf{a},\mathbf{b})$ is to draw a right-angled triangle with $\mathbf{a}$ and $\mathbf{b}$ at its two non-right-angled vertices, as shown in Figure 1.2. Then $e_1(\mathbf{a},\mathbf{b})$ is the sum of the distances along the sides parallel to the axes. Another way to interpret it is to consider $\mathbf{a}$ and $\mathbf{b}$ as representing intersections in a city where the roads form a rectangular grid, as Figure 1.3 illustrates. Then $e_1(\mathbf{a},\mathbf{b})$ is the shortest distance that can be travelled by a vehicle, such as a taxi, to get from $\mathbf{a}$ to $\mathbf{b}$. This interpretation leads to the common name **taxicab metric** for e_1.

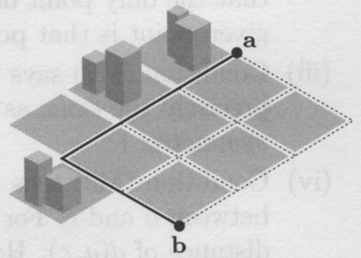


Figure 1.3

Let us now verify that e_1 does define a metric. We check that it satisfies conditions (M1)–(M3).

(M1) For all $\mathbf{a},\mathbf{b}\in\mathbb{R}^2$, $|b_1-a_1|$ and $|b_2-a_2|$ are non-negative, and hence so is their sum: thus $e_1(\mathbf{a},\mathbf{b})\ge 0$.

For all $\mathbf{a}\in\mathbb{R}^2$,

$$e_1(\mathbf{a},\mathbf{a}) = |a_1-a_1|+|a_2-a_2| = 0+0=0.$$

Conversely, suppose $\mathbf{a},\mathbf{b}\in\mathbb{R}^2$ are such that $e_1(\mathbf{a},\mathbf{b})=0$. Then

$$0=|b_1-a_1|+|b_2-a_2|,$$

which implies that both $|b_1-a_1|=0$ and $|b_2-a_2|=0$. Thus $a_1=b_1$ and $a_2=b_2$; that is, $\mathbf{a}=\mathbf{b}$.

This proves that condition (M1) is satisfied.

(M2) Let $\mathbf{a},\mathbf{b}\in\mathbb{R}^2$. Using property (M2) of the modulus function, we obtain

The modulus function is the Euclidean distance function on $\mathbb{R}$, and so is a metric.

$$e_1(\mathbf{a},\mathbf{b}) = |b_1-a_1|+|b_2-a_2| = |a_1-b_1|+|a_2-b_2| = e_1(\mathbf{b},\mathbf{a}).$$

This proves that condition (M2) is satisfied.

(M3) Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^2$. Then, using property (M3) of the modulus function,

$$\begin{aligned} e_1(\mathbf{a}, \mathbf{c}) &= |c_1 - a_1| + |c_2 - a_2| \\ &\leq (|c_1 - b_1| + |b_1 - a_1|) + (|c_2 - b_2| + |b_2 - a_2|) \\ &= (|b_1 - a_1| + |b_2 - a_2|) + (|c_1 - b_1| + |c_2 - b_2|) \\ &= e_1(\mathbf{a}, \mathbf{b}) + e_1(\mathbf{b}, \mathbf{c}). \end{aligned}$$

This proves that condition (M3) is satisfied.

We conclude that e_1 is a metric on the plane.

1.2 The definition of continuity

In *Unit A1*, we discussed the ε–δ definition of continuity of a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. We now generalize this to a definition of continuity for functions between *any* pair of metric spaces.

First, we recall the ε–δ definition of continuity.

The definition in *Unit A1* involves a subset A of $\mathbb{R}^n$ rather than $\mathbb{R}^n$ itself, but we do not need that generalization here.

A function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is **continuous** at $\mathbf{a} \in \mathbb{R}^n$ if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all $\mathbf{x} \in \mathbb{R}^n$,

$$d^{(m)}(f(\mathbf{x}), f(\mathbf{a})) < \varepsilon \quad \text{whenever} \quad d^{(n)}(\mathbf{x}, \mathbf{a}) < \delta.$$

The most important feature of this definition is that the only properties of the domain and codomain that appear are their metrics $d^{(n)}$ and $d^{(m)}$. In consequence, this definition can be adapted to a function between any two metric spaces, as follows.

Definition

Let (X, d) and (Y, e) be metric spaces.

A function $f\colon X \to Y$ is **continuous** at $a \in X$ if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all $x \in X$,

$$e(f(x), f(a)) < \varepsilon \quad \text{whenever} \quad d(x, a) < \delta.$$

A function that is continuous at all points of X is **continuous** on X (or simply **continuous**, if no ambiguity is possible).

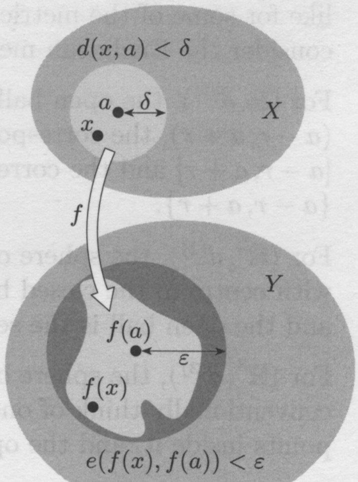


Figure 1.4

Remarks

(i) When we wish to emphasize the particular metrics d and e on X and Y, we say that f is *(d, e)-continuous* at a or on X (or is simply *(d, e)-continuous*).

(ii) We shall refer to this definition of continuity on metric spaces as the **ε–δ definition.**

(iii) When (X, d) and (Y, e) are both Euclidean spaces, we recover the definition of continuity in *Unit A1*.

(iv) We have not defined continuity on a subset A of the metric space X; we return to this issue later.

Subsection 3.1.

If we look at the definition of continuity, we see that a key role is played by those points that are less than a distance δ from a and by those that are less than a distance ε from $f(a)$. There is standard terminology for sets like this.

Definition

Let (X, d) be a metric space, and let $a \in X$ and $r \geq 0$.

The **open ball** of **radius** r with **centre** a is the set

$$B_d(a, r) = \{x : d(a, x) < r\}.$$

The **closed ball** of **radius** r with **centre** a is the set

$$B_d[a, r] = \{x : d(a, x) \leq r\}.$$

The **sphere** of **radius** r with **centre** a is the set

$$S_d(a, r) = \{x : d(a, x) = r\}.$$

When $r = 1$, these sets are called respectively the **unit open ball** with **centre** a, the **unit closed ball** with **centre** a and the **unit sphere** with **centre** a.

There is no universal agreement on the terminology and notation for balls. It is a wise precaution to check the conventions whenever reading a book on metric spaces.

Remark

When we wish to emphasize the particular metric d on X, we refer to the *d-open ball*, the *d-closed ball* and the *d-sphere*.

Problem 1.2

Let (X, d) be a metric space, and let $a \in X$. Show that

$$B_d(a, 0) = \varnothing, \quad B_d[a, 0] = \{a\} \quad \text{and} \quad S_d(a, 0) = \{a\}.$$

It will be helpful to discover what open balls, closed balls and spheres look like for some of the metric spaces we have come across so far. First, we consider the Euclidean metric spaces $(\mathbb{R}, d^{(1)})$, $(\mathbb{R}^2, d^{(2)})$ and $(\mathbb{R}^3, d^{(3)})$.

For $(\mathbb{R}, d^{(1)})$, the open ball of radius r with centre a is the open interval $(a - r, a + r)$, the corresponding closed ball is the closed interval $[a - r, a + r]$ and the corresponding sphere is the set of two points $\{a - r, a + r\}$.

For $(\mathbb{R}^2, d^{(2)})$, the sphere of radius r with centre $\mathbf{a}$ is the circle of radius r with centre $\mathbf{a}$, the closed ball is the circle together with all points inside it, and the open ball is the set of points inside the circle — an open disc.

For $(\mathbb{R}^3, d^{(3)})$, the sphere of radius r with centre $\mathbf{a}$ is a hollow sphere as we conventionally think of one, the closed ball is the sphere together with all points inside it, and the open ball is the set of points inside the sphere.

Next, let us look at open balls defined using the taxicab metric.

Worked problem 1.1

Consider the metric space $(\mathbb{R}^2, e_1)$ — that is, the plane with the taxicab metric. Find the unit open ball $B_{e_1}(\mathbf{0}, 1)$.

Solution

The centre is $\mathbf{0} = (0,0)$, and we want to find all points $\mathbf{x} = (x_1, x_2)$ that satisfy

$$e_1(\mathbf{0}, \mathbf{x}) = |x_1| + |x_2| < 1.$$

We first consider points in the first quadrant, where $x_1, x_2 \geq 0$. We want to find those points where $x_1 + x_2 < 1$. Consider the line $x_1 + x_2 = 1$, or equivalently $x_2 = 1 - x_1$. In the first quadrant, this line connects the points $(0,1)$ and $(1,0)$ and is shown dashed in Figure 1.5. The points on this line segment have coordinates $(x_1, 1 - x_1)$. All points below the line segment have coordinates (x_1, x_2) with $x_2 < 1 - x_1$ and all points on or above it have coordinates (x_1, x_2) with $x_2 \geq 1 - x_1$. Hence the points where $x_1 + x_2 < 1$ are those strictly below the line segment, comprising the shaded region in Figure 1.6.

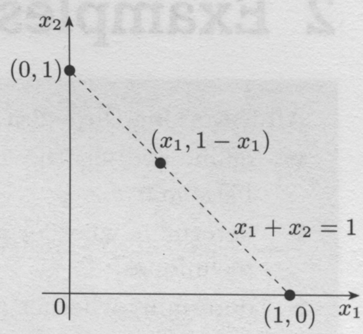


Figure 1.5

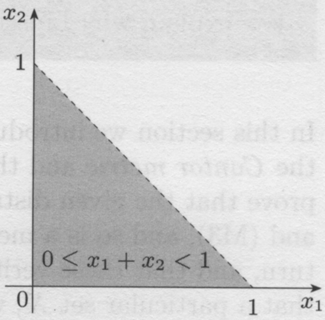


Figure 1.6

By use of a similar argument for each of the other three quadrants, or by appealing to the symmetry of the situation, we obtain triangular regions in each quadrant. Combining these we obtain the diamond-shaped region in Figure 1.7, and the open ball $B_{e_1}(\mathbf{0}, 1)$ is the set of points strictly inside this diamond, shown shaded in the figure. ∎

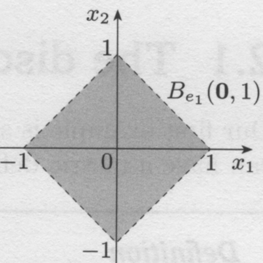


Figure 1.7

Problem 1.3

Sketch the open ball $B_{e_1}((2,3), 2)$, briefly justifying your answer.

We can rephrase the definition of continuity in terms of balls as follows.

> **Definition**
>
> Let (X, d) and (Y, e) be metric spaces.
>
> A function $f: X \to Y$ is **continuous** at $a \in X$ if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that
>
> $$f(x) \in B_e(f(a), \varepsilon) \quad \text{whenever} \quad x \in B_d(a, \delta).$$
>
> A function that is continuous at all points of X is **continuous** on X.

Remarks

(i) We shall refer to this definition of continuity on metric spaces as the **definition in terms of balls**.

(ii) The two definitions of continuity on metric spaces that we have given — the ε–δ definition and the definition in terms of balls — are equivalent. Elsewhere in this unit, and in the course, we shall use whichever definition seems more suited to the purpose at hand.

2 Examples of metrics

After working through this section, you should be able to:

- define and use the *discrete metric*, the *Cantor metric* and the *max metric*;
- determine whether a given function is a metric, using standard techniques;
- determine whether a given function between two metric spaces is continuous;
- state the definition of a *Lipschitz function*, and determine whether a given function is Lipschitz;
- explain why Lipschitz functions are continuous.

In this section we introduce three examples of metrics: the *discrete metric*, the *Cantor metric* and the *max metric*. In each case, our main task is to prove that the given distance function satisfies conditions (M1), (M2) and (M3), and so is a metric. You will see that each condition is verified in turn, and that these verifications follow a similar pattern. Once we know that a particular set X, with distance function d, *does* form a metric space (X, d), we look at examples of continuous functions defined on it.

As you will see, the discrete metric can be defined on any set whatsoever, whereas the others are defined on particular sets.

The section ends by looking briefly at *Lipschitz functions.*

2.1 The discrete metric

Our first example is simple but important, since it shows that *every* set can have a metric defined on it.

Definition

Let X be a set. The **discrete metric** on X is the function $d_0 : X \times X \to \mathbb{R}$ defined by

$$d_0(x,y) = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y. \end{cases}$$

We show that d_0 is a metric after the remarks below.

Remarks

(i) When $X = \varnothing$, then the definition still makes sense but is not very interesting.

(ii) The d_0-distance between any two *distinct* points of a set X is always equal to 1. In particular, when $X = \mathbb{R}$, since $d_0(0,2) = 1$ and $d^{(1)}(0,2) = 2$, the discrete metric gives a second metric on $\mathbb{R}$. In fact, if (X,d) is any metric space whose metric d is not d_0, then (X, d_0) is a second metric space with underlying set X.

Theorem 2.1

If X is any set, then (X, d_0) is a metric space.

Proof In order to check that d_0 is a metric on X, we show that d_0 satisfies conditions (M1)–(M3).

If $X = \varnothing$, then X contains no elements and there is nothing to verify.

Now suppose that $X \neq \varnothing$.

(M1) Since d_0 can take only the values 0 and 1, we have $d_0(a,b) \geq 0$ for all $a, b \in X$.

The definition of d_0 implies immediately that $d_0(a,b) = 0$ if and only if $a = b$.

Thus d_0 satisfies (M1).

(M2) If $a = b$, then $d_0(a,b) = 0 = d_0(b,a)$.

If $a \neq b$, then $d_0(a,b) = 1 = d_0(b,a)$.

Thus, for all $a, b \in X$, (M2) holds.

(M3) Let $a, b, c \in X$. We examine the two possible cases: $d_0(a,c) = 0$ and $d_0(a,c) = 1$.

Suppose $d_0(a,c) = 0$ (so $a = c$). Since $d_0(a,b)$ and $d_0(b,c)$ are non-negative, it follows that

$$d_0(a,b) + d_0(b,c) \geq 0 = d_0(a,c).$$

Now suppose $d_0(a,c) = 1$; then $a \neq c$ and so b cannot equal both a and c. Hence, from (M1), at least one of $d_0(a,b)$ and $d_0(b,c)$ is non-zero and so must equal 1. Thus

$$d_0(a,b) + d_0(b,c) \geq 1 = d_0(a,c).$$

Thus, in both cases, (M3) holds.

Since d_0 satisfies (M1)–(M3), it is a metric on X. Hence (X, d_0) is a metric space. ■

Remark

At first appearance, the discrete metric d_0 may not seem important. Its utility lies in the fact that it gives an 'extreme' example of a metric space — no other definition of distance so completely ignores any structure that may be present in the set X. We use the discrete metric for a number of purposes, principally to test properties that we suspect may hold for all metric spaces. If such a property fails for (X, d_0), then we know that our suspicion was false.

Problem 2.1

Let $X = \{x, y, z\}$ and define $d : X \times X \to \mathbb{R}$ by

$$d(x,x) = d(y,y) = d(z,z) = 0;$$
$$d(x,y) = d(y,x) = 1;$$
$$d(y,z) = d(z,y) = 2;$$
$$d(x,z) = d(z,x) = 4.$$

Determine whether d is a metric on X.

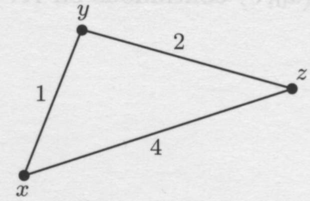


Figure 2.1

Continuity and the discrete metric

Suppose that (X, d_0) and (Y, e) are metric spaces and that $f: X \to Y$ is a function between them. What can we say about the continuity of f at a point $a \in X$?

Note that X has the discrete metric.

Using the definition of continuity in terms of balls, we see that f is continuous at a if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$f(x) \in B_e(f(a), \varepsilon) \quad \text{whenever} \quad x \in B_{d_0}(a, \delta).$$

Let us look at the form of $B_{d_0}(a, \delta)$ for various $\delta > 0$.

There are two cases to consider.

- ▶ If $0 < \delta \leq 1$ and $d_0(a, x) < \delta$, then $d_0(a, x) = 0$ and so $x = a$. Thus,

  $$B_{d_0}(a, \delta) = \{x \in X : d_0(a, x) < \delta\} = \{a\}.$$

- ▶ If $\delta > 1$, then $d_0(a, x) \leq 1 < \delta$ for all $x \in X$. Thus,

  $$B_{d_0}(a, \delta) = \{x \in X : d_0(a, x) < \delta\} = X.$$

This has an interesting consequence for the continuity of an arbitrary function f between (X, d_0) and (Y, e).

> ***Theorem 2.2***
>
> Let (Y, e) be a metric space, let X be any set and let $f: X \to Y$.
>
> Then f is (d_0, e)-continuous on X.

By Theorem 2.1, we can attach the metric d_0 to any set we please.

Proof Let $a \in X$ and consider (d_0, e)-continuity at a.

We must show that, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$f(x) \in B_e(f(a), \varepsilon) \quad \text{whenever} \quad x \in B_{d_0}(a, \delta).$$

Let $\varepsilon > 0$ be given.

We have just seen that if $0 < \delta \leq 1$ then

$$B_{d_0}(a, \delta) = \{a\}.$$

Since $f(a) \in B_e(f(a), \varepsilon)$, it follows that, for such a δ,

$$f(B_{d_0}(a, \delta)) = \{f(a)\} \subseteq B_e(f(a), \varepsilon).$$

In particular, if we set $\delta = \frac{1}{2}$, then, for each $\varepsilon > 0$,

$$f(x) \in B_e(f(a), \varepsilon) \quad \text{whenever} \quad x \in B_{d_0}(a, \delta).$$

Thus f is (d_0, e)-continuous at a. Since a is an arbitrary point of X, f is (d_0, e)-continuous on X. ■

Remarks

(i) Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} -1 & \text{if } x \in \mathbb{Q}, \\ 1 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

As a mapping from $(\mathbb{R}, d^{(1)})$ to $(\mathbb{R}, d^{(1)})$, f is *discontinuous everywhere*: this accords with our intuition. However, Theorem 2.2 tells us that, as a mapping from $(\mathbb{R}, d_0)$ to $(\mathbb{R}, d^{(1)})$, f is *continuous everywhere*! Thus:

continuity depends on how we measure distance.

(ii) The situation when the discrete metric is applied to the codomain Y rather than the domain X of f is more complicated and is related to the idea of *connectedness*. We defer discussion of this until *Unit C1*.

2.2 The Cantor metric

In this subsection, we describe a metric defined on a certain class of real sequences. The resulting metric space belongs to a class of metric spaces known as *sequence spaces*. These spaces have applications in many areas of mathematics.

Definition

The **Cantor space C** consists of all *infinite* sequences of zeroes and ones:

$$\mathbf{C} = \{(a_n) : a_n \in \{0, 1\} \text{ for all } n \in \mathbb{N}\}.$$

Georg Cantor (1845–1918) is remembered for his fundamental contributions to set theory — in particular, for his astonishing result that there are different 'sizes' of infinite set: for example, $\mathbb{R}$ is 'bigger' than $\mathbb{Z}$ or $\mathbb{Q}$.

Remarks

(i) For ease of writing, we denote points in $\mathbf{C}$ by bold letters: so $\mathbf{x} = (x_n)$ denotes the point $(x_1, x_2, x_3, \ldots)$.

(ii) Two points in $\mathbf{C}$ are the same if and only if all terms agree: that is, $\mathbf{x} = \mathbf{y}$ if and only if $x_n = y_n$ for all $n \in \mathbb{N}$.

Examples of points in $\mathbf{C}$ are

$$(1, 1, 1, 1, \ldots, 1, \ldots), \quad (0, 0, 0, 0, \ldots, 0, \ldots), \quad (1, 0, 1, 0, \ldots, 1, 0, \ldots).$$

Suppose we are given two points $\mathbf{x} = (x_n)$ and $\mathbf{y} = (y_n)$ in $\mathbf{C}$. How can we measure their distance apart? One idea would be to count the number of places (or terms) where $\mathbf{x}$ and $\mathbf{y}$ differ. Unfortunately, $\mathbf{x}$ and $\mathbf{y}$ can differ at *infinitely* many terms, and so this will not work; points in a space are not allowed to be infinitely far apart!

Consider $(0, 0, 0, 0, \ldots)$ and $(1, 1, 1, 1, \ldots)$, for example.

There are various ways to resolve this problem. One way is to find the first term at which the two sequences differ, and use this as a way of defining the distance between them. Intuitively, such a definition of distance should give smaller values the longer it takes to find a term at which the sequences differ.

With this in mind, consider the following four sequences:

$$\mathbf{a} = (0,0,0,0,0,\ldots); \quad \mathbf{b} = (1,1,0,0,0,\ldots);$$
$$\mathbf{c} = (1,0,0,0,0,\ldots); \quad \mathbf{d} = (1,0,1,1,1,\ldots).$$

The sequence **a** first differs from **b** at the first term; we could say that they are a distance $2^{-1} = \frac{1}{2}$ apart. Next, **b** and **c** first differ at the second term; we could say that they are a distance $2^{-2} = \frac{1}{4}$ apart. Finally, **c** and **d** first differ at the third term; we could say that they are a distance $2^{-3} = \frac{1}{8}$ apart. In this way we get smaller distances the further along the sequences we have to go.

This leads to the following definition of distance in **C**.

> ***Definition***
>
> Let $\mathbf{x}, \mathbf{y} \in \mathbf{C}$. The **Cantor distance** $d_{\mathbf{C}} : \mathbf{C} \times \mathbf{C} \to \mathbb{R}$ between **x** and **y** is
>
> $$d_{\mathbf{C}}(\mathbf{x},\mathbf{y}) = \begin{cases} 0 & \text{if } \mathbf{x} = \mathbf{y}, \\ 2^{-n} & \text{if } \mathbf{x} \text{ and } \mathbf{y} \text{ first differ at the } n\text{th term.} \end{cases}$$

Note that the Cantor distance between *any* two points of **C** is at most $\frac{1}{2}$.

Problem 2.2

Find $d_{\mathbf{C}}(\mathbf{a},\mathbf{e})$, $d_{\mathbf{C}}(\mathbf{b},\mathbf{e})$, $d_{\mathbf{C}}(\mathbf{c},\mathbf{e})$ and $d_{\mathbf{C}}(\mathbf{d},\mathbf{e})$ when

$$\mathbf{a} = (1,0,0,0,0,0,0,\ldots), \quad \mathbf{b} = (1,1,0,0,0,0,0,\ldots),$$
$$\mathbf{c} = (1,1,1,0,0,0,0,\ldots), \quad \mathbf{d} = (1,1,1,1,0,0,0,\ldots),$$
$$\mathbf{e} = (1,1,1,1,1,1,1,\ldots).$$

We now show that $d_{\mathbf{C}}$ is a metric on **C**.

> ***Theorem 2.3***
>
> The Cantor distance is a metric on **C**.

Proof We verify that $d_{\mathbf{C}}$ satisfies conditions (M1)–(M3).

(M1) For all $\mathbf{a}, \mathbf{b} \in \mathbf{C}$, by definition, $d_{\mathbf{C}}(\mathbf{a},\mathbf{b}) \geq 0$.

Also, $d_{\mathbf{C}}(\mathbf{a},\mathbf{a}) = 0$.

If $d_{\mathbf{C}}(\mathbf{a},\mathbf{b}) = 0$, then, by definition, $\mathbf{a} = \mathbf{b}$.

Hence $d_{\mathbf{C}}$ satisfies (M1).

(M2) Let $\mathbf{a}, \mathbf{b} \in \mathbf{C}$. If $\mathbf{a} = \mathbf{b}$, then, by definition,

$$d_{\mathbf{C}}(\mathbf{a},\mathbf{b}) = d_{\mathbf{C}}(\mathbf{b},\mathbf{a}) = 0.$$

If $\mathbf{a} \neq \mathbf{b}$ and **a** and **b** first differ at the nth term, then **b** and **a** also first differ at the nth term. Hence,

$$d_{\mathbf{C}}(\mathbf{a},\mathbf{b}) = d_{\mathbf{C}}(\mathbf{b},\mathbf{a}) = 2^{-n}.$$

Thus, (M2) holds.

(M3) Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbf{C}$.

If $\mathbf{a} = \mathbf{b}$, so that $d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = 0$, then

$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{c}) = d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}) \leq d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) + d_{\mathbf{C}}(\mathbf{b}, \mathbf{c})$$

and so (M3) holds.

A similar argument applies if $\mathbf{b} = \mathbf{c}$, while if $\mathbf{a} = \mathbf{c}$ then $d_{\mathbf{C}}(\mathbf{a}, \mathbf{c}) = 0$, and once again (M3) holds.

So suppose that $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are all distinct, with $\mathbf{a}$ first differing from $\mathbf{b}$ at the mth term and $\mathbf{a}$ first differing from $\mathbf{c}$ at the nth term. There are two possible cases.

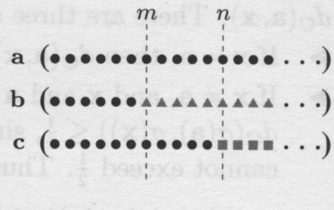


Figure 2.2

If $m \leq n$, as shown in Figure 2.2, then

$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{c}) = 2^{-n} \leq 2^{-m} = d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) \leq d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) + d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}).$$

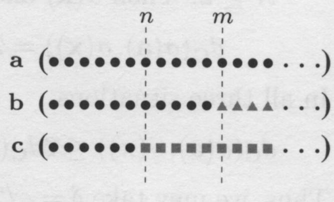
If $m > n$, as shown in Figure 2.3, then $\mathbf{b}$ and $\mathbf{c}$ must differ first at the nth term (since $\mathbf{a}$ agrees with both of them for the first $n-1$ terms and agrees with $\mathbf{b}$ but not with $\mathbf{c}$ at the nth term). Thus,

$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{c}) = d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}).$$

It follows that

$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{c}) = 2^{-n} = d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}) \leq d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) + d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}).$$

Thus, in both cases, (M3) holds.



Since $d_{\mathbf{C}}$ satisfies (M1)–(M3), it is a metric on $\mathbf{C}$. ■

Figure 2.3

Remark

Since the Cantor distance is a metric, it is often referred to as the **Cantor metric**.

There are many other examples of sequence spaces in mathematics. Some of the most useful involve sequences of real numbers with specific properties. The method we have given here serves as a model for defining a metric in these more sophisticated settings.

In the problems for this unit, we ask you to show that some other sequence spaces are metric spaces.

Continuity and the Cantor metric

The Cantor space occurs in many different areas of mathematics and there are many examples of continuous functions defined upon it. One simple example is the following.

> ***Definition***
>
> The **shift map** is the map $\sigma: \mathbf{C} \to \mathbf{C}$ given by
>
> $$\sigma((x_1, x_2, x_3, x_4, \ldots)) = (x_2, x_3, x_4, x_5, \ldots).$$

The shift map 'shifts' the sequence by one term. Such maps play an important role in the theory of dynamical systems.

Thus the shift map acts by deleting the first term of the given sequence: for example,

$$\sigma(0, 1, 1, 1, 1, 1, \ldots) = (1, 1, 1, 1, 1, \ldots),$$
$$\sigma(0, 1, 0, 1, 0, 1, \ldots) = (1, 0, 1, 0, 1, \ldots),$$
$$\sigma(1, 1, 1, 1, 1, 1, \ldots) = (1, 1, 1, 1, 1, \ldots).$$

Worked problem 2.1

Show that the shift map is $(d_{\mathbf{C}}, d_{\mathbf{C}})$-continuous on $\mathbf{C}$.

Solution

Let $\mathbf{a} = (a_n) \in \mathbf{C}$ and let $\varepsilon > 0$ be given. We must find a $\delta > 0$ such that

$$d_{\mathbf{C}}(\sigma(\mathbf{a}), \sigma(\mathbf{x})) < \varepsilon \quad \text{whenever} \quad d_{\mathbf{C}}(\mathbf{a}, \mathbf{x}) < \delta.$$

Here, and in the remainder of this section, we use the ε–δ definition of continuity rather than the definition in terms of balls.

To do this, we need to find an upper bound for $d_{\mathbf{C}}(\sigma(\mathbf{a}), \sigma(\mathbf{x}))$ in terms of $d_{\mathbf{C}}(\mathbf{a}, \mathbf{x})$. There are three situations to consider.

- If $\mathbf{x} = \mathbf{a}$, then $d_{\mathbf{C}}(\mathbf{a}, \mathbf{x}) = 0$ and $d_{\mathbf{C}}(\sigma(\mathbf{a}), \sigma(\mathbf{x})) = 0$.
- If $\mathbf{x} \neq \mathbf{a}$, and $\mathbf{x}$ and $\mathbf{a}$ differ at the first term, then $d_{\mathbf{C}}(\mathbf{a}, \mathbf{x}) = \frac{1}{2}$ and $d_{\mathbf{C}}(\sigma(\mathbf{a}), \sigma(\mathbf{x})) \leq \frac{1}{2}$, since the Cantor distance between two points of $\mathbf{C}$ cannot exceed $\frac{1}{2}$. Thus,
$$d_{\mathbf{C}}(\sigma(\mathbf{a}), \sigma(\mathbf{x})) \leq \tfrac{1}{2} = d_{\mathbf{C}}(\mathbf{a}, \mathbf{x}).$$
- If $\mathbf{x} \neq \mathbf{a}$, but their first terms are equal, then they must *first* differ at some later term. Suppose that they first differ at the Nth term, where $N \geq 2$. Then $\sigma(\mathbf{x})$ and $\sigma(\mathbf{a})$ first differ at the $(N-1)$th term, and so
$$d_{\mathbf{C}}(\sigma(\mathbf{a}), \sigma(\mathbf{x})) = 2^{-(N-1)} = 2 \times 2^{-N} = 2d_{\mathbf{C}}(\mathbf{a}, \mathbf{x}).$$

In all three situations,

$$d_{\mathbf{C}}(\sigma(\mathbf{a}), \sigma(\mathbf{x})) \leq 2d_{\mathbf{C}}(\mathbf{a}, \mathbf{x}).$$

Thus, we may take $\delta = \varepsilon/2$; then $d_{\mathbf{C}}(\sigma(\mathbf{a}), \sigma(\mathbf{x})) < \varepsilon$ whenever $d_{\mathbf{C}}(\mathbf{a}, \mathbf{x}) < \delta$.

It follows that σ is $(d_{\mathbf{C}}, d_{\mathbf{C}})$-continuous at $\mathbf{a}$. Since $\mathbf{a}$ is an arbitrary point of $\mathbf{C}$, it follows that σ is $(d_{\mathbf{C}}, d_{\mathbf{C}})$-continuous on $\mathbf{C}$. ■

2.3 The max metric on $C[0, 1]$

Our next example is of a metric defined on a set of functions — that is, we consider 'distance between functions'. The functions in question are the continuous functions from the closed interval $[0, 1]$ to $\mathbb{R}$, and the metric is known as the *max metric*.

> ***Notation***
>
> $C[0, 1]$ denotes the set of all continuous functions $f\colon [0, 1] \to \mathbb{R}$.

Remarks

(i) The points of $C[0, 1]$ are continuous functions on $[0, 1]$, such as $f(x) = 3x$ and $g(x) = \sin x$, for $x \in [0, 1]$.

(ii) In a similar way, we can define the set $C[a, b]$ of all continuous functions from the closed interval $[a, b]$ to $\mathbb{R}$.

In discussing the max matric, we are going to need the following properties of the functions in the set $C[0, 1]$, all of which can be deduced from results given in *Unit A1*.

These properties can be deduced from the Combination Rules for continuous functions from $\mathbb{R}$ to $\mathbb{R}$ and the result of Problem 2.5 of *Unit A1*, together with the Restriction Rule (Theorem 2.1 of *Unit A1*).

Properties of functions in $C[0,1]$

Let f and g be real-valued continuous functions with domain $[0,1]$, i.e. $f, g \in C[0,1]$, and let $\lambda \in \mathbb{R}$. Then the following real-valued functions are all continuous on $[0,1]$, i.e. they are all points in $C[0,1]$:

Sum $\quad f+g\colon[0,1] \to \mathbb{R}$, defined by $(f+g)(x) = f(x) + g(x)$;

Multiple $\quad \lambda f\colon [0,1] \to \mathbb{R}$, defined by $(\lambda f)(x) = \lambda \times f(x)$;

Product $\quad fg\colon [0,1] \to \mathbb{R}$, defined by $(fg)(x) = f(x)g(x)$;

Modulus $\quad |f|\colon [0,1] \to \mathbb{R}$, defined by $|f|(x) = |f(x)|$.

A *real-valued function* is one with codomain $\mathbb{R}$.

The max metric on $C[0,1]$ is based on the idea of measuring the distance between two functions f and g in $C[0,1]$ by the maximum of their difference over the interval $[0,1]$. This is given by

$$\max\{|g(x) - f(x)| : x \in [0,1]\}$$

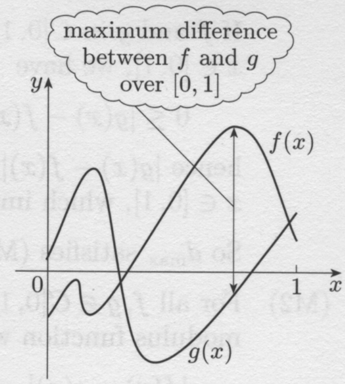


Figure 2.4

and is illustrated in Figure 2.4. It makes sense to talk of this quantity, because the map $x \mapsto |g(x) - f(x)|$ defines a continuous function on $[0,1]$ and hence, by the Extreme Value Theorem, this maximum exists. We thus make the following definition.

Unit A1, Theorem 2.5.

Definition

The **max metric** on $C[0,1]$ is the function $d_{\max}\colon C[0,1] \times C[0,1] \to \mathbb{R}$ defined by

$$d_{\max}(f,g) = \max\{|g(x) - f(x)| : x \in [0,1]\}.$$

We show that $d_{\max}$ is a metric in the next theorem.

Problem 2.3

Find $d_{\max}(f,g)$ when $f, g\colon [0,1] \to \mathbb{R}$ are given by:

(a) $f(x) = 1$, $g(x) = x$;

(b) $f(x) = \frac{1}{2}$, $g(x) = \sin 2\pi x$;

(c) $f(x) = x$, $g(x) = x^2$.

Hint In answering (c), you might find it helpful to write $h = g - f$ and to recall that the maximum and minimum of a function h can be found by considering $h(x)$ at the points x where $h'(x) = 0$ and at the endpoints $x = 0$, $x = 1$. You may also find it helpful to sketch the graph of h.

We now show that $d_{\max}$ is indeed a metric on $C[0,1]$.

Theorem 2.4

$(C[0,1], d_{\max})$ is a metric space.

Proof We show that $d_{\max}$ satisfies conditions (M1)–(M3).

(M1) Since $|g(x) - f(x)| \geq 0$ for all $x \in [0,1]$, its maximum is also non-negative. So $d_{\max}(f,g) \geq 0$ for all f and g in $C[0,1]$.

For all $f \in C[0,1]$ and all $x \in [0,1]$, we have $|f(x) - f(x)| = 0$; therefore $d_{\max}(f,f) = 0$ for $f \in C[0,1]$.

If f and g in $C[0,1]$ are such that $d_{\max}(f,g) = 0$, then, for all $x \in [0,1]$, we have

$$0 \leq |g(x) - f(x)| \leq d_{\max}(f,g) = 0;$$

hence $|g(x) - f(x)| = 0$ for all $x \in [0,1]$. Thus, $f(x) = g(x)$ for all $x \in [0,1]$, which implies that $f = g$.

So $d_{\max}$ satisfies (M1).

(M2) For all $f, g \in C[0,1]$ and all $x \in [0,1]$, using property (M2) of the modulus function we have

$$|f(x) - g(x)| = |g(x) - f(x)|,$$

and so

$$\begin{aligned} d_{\max}(g,f) &= \max\{|f(x) - g(x)| : x \in [0,1]\} \\ &= \max\{|g(x) - f(x)| : x \in [0,1]\} = d_{\max}(f,g). \end{aligned}$$

So $d_{\max}$ satisfies (M2).

(M3) Let $f, g, h \in C[0,1]$ and $x \in [0,1]$. By property (M3) of the modulus function, we have

$$|h(x) - f(x)| \leq |h(x) - g(x)| + |g(x) - f(x)|. \tag{2.1}$$

Now, for all $x \in [0,1]$,

$$|h(x) - g(x)| \leq \max\{|h(x) - g(x)| : x \in [0,1]\} = d_{\max}(g,h), \tag{2.2}$$

and similarly,

$$|g(x) - f(x)| \leq d_{\max}(f,g). \tag{2.3}$$

Thus, for each $x \in [0,1]$, combining (2.1), (2.2) and (2.3) we have

$$|h(x) - f(x)| \leq d_{\max}(f,g) + d_{\max}(g,h).$$

Since this holds for each $|h(x) - f(x)|$ with $x \in [0,1]$, it is true for the maximum value, and so

$$d_{\max}(f,h) \leq d_{\max}(f,g) + d_{\max}(g,h).$$

Thus, $d_{\max}$ satisfies (M3).

Since $d_{\max}$ satisfies conditions (M1)–(M3), it is a metric on $C[0,1]$. Hence $(C[0,1], d_{\max})$ is a metric space. ■

Remark

Similarly, for any interval $[a,b]$, we can define a metric $d_{\max}$ on $C[a,b]$ by

$$d_{\max}(f,g) = \max\{|g(x) - f(x)| : x \in [a,b]\}.$$

Problem 2.4

For $f, g \in C[0,1]$, define $d_{\min}(f,g)$ by

$$d_{\min}(f,g) = \min\{|g(x) - f(x)| : x \in [0,1]\}.$$

Determine whether $d_{\min}$ is a metric on $C[0,1]$.

Continuity and the max metric on $C[0,1]$

One way to learn about the nature of a metric space is to consider functions from the space to itself and ask whether they are continuous. But how do we define a function from $(C[0,1], d_{\max})$ to itself? A simple way is as follows.

Let us add a constant real number c to a function $f \in C[0,1]$; then the new function, given by $f_c(x) = f(x) + c$ for $x \in [0,1]$, is also continuous. Thus we may define a map $F_c : C[0,1] \to C[0,1]$ by $f \mapsto f_c$ for $f \in C[0,1]$.

Problem 2.5

Write down $F_c(f)$ when:

(a) $c = 1$ and $f(x) = x^2$;

(b) $c = \pi$ and $f(x) = \sin x$.

Worked problem 2.2

Show that F_c is $(d_{\max}, d_{\max})$-continuous on $C[0,1]$, for any $c \in \mathbb{R}$.

Solution

Let $c \in \mathbb{R}$. Let $f \in C[0,1]$ and consider the continuity of F_c at f.

Let $\varepsilon > 0$ be given. We must find a $\delta > 0$ such that, for all $g \in C[0,1]$,

$$d_{\max}(F_c(f), F_c(g)) < \varepsilon \quad \text{whenever} \quad d_{\max}(f,g) < \delta.$$

We first find an upper bound for $d_{\max}(F_c(f), F_c(g))$ in terms of $d_{\max}(f,g)$. We know that

$$d_{\max}(F_c(f), F_c(g)) = \max\{|F_c(g)(x) - F_c(f)(x)| : x \in [0,1]\}.$$

Let us consider the value of $|F_c(g)(x) - F_c(f)(x)|$ for a *particular* $x \in [0,1]$. We find that

This is a common technique for dealing with situations involving max.

$$|F_c(g)(x) - F_c(f)(x)| = |(g(x)+c) - (f(x)+c)| = |g(x) - f(x)|.$$

Hence, for any $x \in [0,1]$,

$$|F_c(g)(x) - F_c(f)(x)| = |g(x) - f(x)|$$

and so

$$\max\{|F_c(g)(x) - F_c(f)(x)| : x \in [0,1]\} = \max\{|g(x) - f(x)| : x \in [0,1]\}.$$

That is,

$$d_{\max}(F_c(f), F_c(g)) = d_{\max}(f,g).$$

Now take $\delta = \varepsilon$. Then, whenever $d_{\max}(f,g) < \delta$, we also have

$$d_{\max}(F_c(f), F_c(g)) < \varepsilon,$$

and so F_c is $(d_{\max}, d_{\max})$-continuous at f.

Since f is an arbitrary point of $C[0,1]$ (that is, f is an arbitrary continuous function on [0,1]), F_c is continuous on $C[0,1]$. Since c is an arbitrary point of $\mathbb{R}$, F_c is continuous on $C[0,1]$ for any $c \in \mathbb{R}$. ■

Instead of defining a function from $C[0,1]$ to $C[0,1]$ by adding a fixed *number* to each $f \in C[0,1]$, let us be bolder and add a fixed *function.*

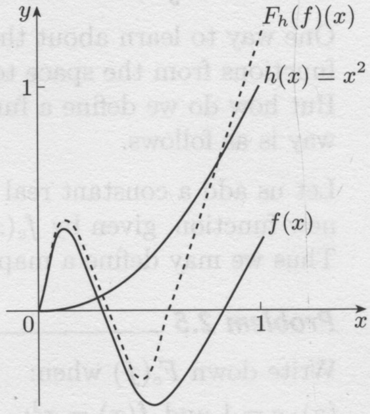


Figure 2.5

Let $h \in C[0,1]$ be given by $h(x) = x^2$, and define the map F_h by $f \mapsto f + h$ for $f \in C[0,1]$. Then, for each $f \in C[0,1]$, the function $F_h(f)$ is a function on $[0,1]$ whose value at a particular point $x \in [0,1]$ is given by

$$F_h(f)(x) = f(x) + x^2.$$

Since f and h are in $C[0,1]$, $f + h$ is also in $C[0,1]$ and so $F_h: C[0,1] \to C[0,1]$.

The proof that $F_c: C[0,1] \to C[0,1]$ is continuous for any constant $c \in \mathbb{R}$ extends quite easily to a proof that $F_h: C[0,1] \to C[0,1]$ is continuous for any function $h \in C[0,1]$. We now ask you to show this.

Problem 2.6

Show that $F_h: C[0,1] \to C[0,1]$ is $(d_{\max}, d_{\max})$-continuous on $C[0,1]$, for any $h \in C[0,1]$.

Worked problem 2.2 can be thought of as a special case of Problem 2.6 with h defined by $h(x) = c$ for all $x \in [0,1]$.

We have seen that adding a fixed number or function to a function in $C[0,1]$ produces a continuous function from $(C[0,1], d_{\max})$ to itself. Similarly, multiplying by a fixed number or function produces a continuous function from $(C[0,1], d_{\max})$ to itself. The case where we multiply by a number is a special case of multiplying by a function, so let us just consider the latter case. We choose a function $h \in C[0,1]$ and define the *product* function by $G_h(f) = hf$, for $f \in C[0,1]$. Since f and h are in $C[0,1]$, hf is also in $C[0,1]$ and so $G_h: C[0,1] \to C[0,1]$.

Thus $G_h(f)(x) = h(x) \times f(x)$.

Worked problem 2.3

Show that G_h is $(d_{\max}, d_{\max})$-continuous on $C[0,1]$, for any $h \in C[0,1]$.

Solution

Let $h \in C[0,1]$. Let $f \in C[0,1]$ and consider the continuity of G_h at f.

Let $\varepsilon > 0$ be given. We must find a $\delta > 0$ such that, for all $g \in C[0,1]$,

$$d_{\max}(G_h(f), G_h(g)) < \varepsilon \quad \text{whenever} \quad d_{\max}(f,g) < \delta.$$

We first find an upper bound for $d_{\max}(G_h(f), G_h(g))$ in terms of $d_{\max}(f,g)$. We know that

$$d_{\max}(G_h(f), G_h(g)) = d_{\max}(hf, hg) = \max\{|h(x)g(x) - h(x)f(x)| : x \in [0,1]\}.$$

Now, for all $x \in [0,1]$,

$$\begin{aligned} |h(x)g(x) - h(x)f(x)| &= |h(x)(g(x) - f(x))| \\ &= |h(x)|\,|g(x) - f(x)| \le |h(x)|\, d_{\max}(f,g). \end{aligned}$$

So we must find an upper bound for $h(x)$ in which x does not appear. But $h \in C[0,1]$, and so $|h| \in C[0,1]$. Hence, by the Boundedness Theorem, there is a real number $M > 0$ such that

Unit A1, Theorem 2.4.

$$|h(x)| \le M, \qquad \text{for all } x \in [0,1].$$

It follows that there is a constant $M > 0$ such that, for all $x \in [0,1]$,

$$|h(x)g(x) - h(x)f(x)| \le M d_{\max}(f,g),$$

and so

$$d_{\max}(G_h(f), G_h(g)) \le M d_{\max}(f,g).$$

M must be at least 0 (by the nature of the modulus function), and we may always take M to be positive, since increasing M cannot make the inequality false.

Hence, if $d_{\max}(f,g) < \delta$, then

$$d_{\max}(G_h(f), G_h(g)) < M\delta.$$

Now take $\delta = \varepsilon/M$. Then, whenever $d_{\max}(f,g) < \delta$, we also have

$$d_{\max}(G_h(f), G_h(g)) < \varepsilon.$$

This proves that G_h is continuous at f.

Since f is an arbitrary point of $C[0,1]$, G_h is continuous on $C[0,1]$. Since h is an arbitrary point of $C[0,1]$, G_h is continuous on $C[0,1]$ for any $h \in C[0,1]$. ■

Open and closed balls in the max metric

Given a function $f \in C[0,1]$ and $r > 0$, the open ball $B_{d_{\max}}(f,r)$ is determined by finding the set of all functions $g \in C[0,1]$ satisfying the inequality $d_{\max}(f,g) < r$ — that is,

$$d_{\max}(f,g) = \max\{|g(x) - f(x)| : x \in [0,1]\} < r.$$

For each $x \in [0,1]$, if $|g(x) - f(x)| < r$ then $g(x)$ must lie within a distance r of $f(x)$. Thus, over the whole of $[0,1]$, $g(x)$ must lie within a band following the shape of f and extending r units above and below it at each point, as Figure 2.6 illustrates. It follows that the open ball $B_{d_{\max}}(f,r)$ consists of those $g \in C[0,1]$ whose graphs lie within this band.

The closed ball $B_{d_{\max}}[f,r]$ is obtained from this by extending the band to include the points a distance r above and below the graph of f.

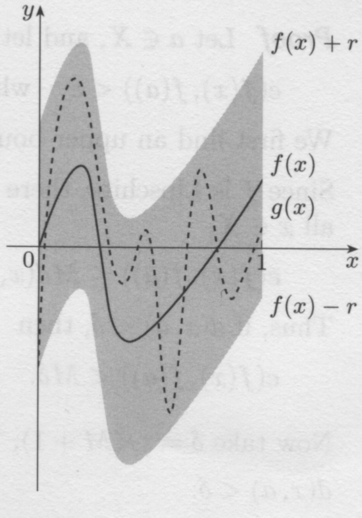


Figure 2.6

2.4 Lipschitz functions

Given two metric spaces, there is a useful class of functions between them whose members are always continuous.

In 1864, the Russian mathematician Rudolf Lipschitz introduced the class of functions $f\colon \mathbb{R} \to \mathbb{R}$ that satisfy the bound $|f(b) - f(a)| \le M|b-a|$, for all $a, b \in \mathbb{R}$, where M is a non-negative constant. Lipschitz functions are important in the theory of differential equations. We shall also find them useful later in this unit.

Definition

Let (X,d) and (Y,e) be metric spaces.

A function $f\colon X \to Y$ is a **Lipschitz function** if there is a non-negative real number M such that, for all $a, b \in X$,

$$e(f(a), f(b)) \le M d(a,b).$$

Remarks

(i) When we wish to emphasize the particular metrics d and e on X and Y, we say that f is a (d,e)-*Lipschitz function*.

(ii) When $(X,d) = (Y,e) = (\mathbb{R}, d^{(1)})$, then the definition reads:

A function $f\colon \mathbb{R} \to \mathbb{R}$ is a Lipschitz function if there is a non-negative real number M such that, for all $a, b \in \mathbb{R}$,

$$|f(b) - f(a)| \le M|b-a|.$$

For example, it can be shown that $f(x) = \sin x$ satisfies

$$|\sin b - \sin a| \le |b - a|$$

for all $a, b \in \mathbb{R}$; thus the sine function is Lipschitz (with $M = 1$).

This follows from the Mean Value Theorem (see the Handbook).

The form of the inequality in the definition is ready-made for proving that Lipschitz functions are continuous.

Theorem 2.5

Let (X, d) and (Y, e) be metric spaces.

If a function $f: X \to Y$ is (d, e)-Lipschitz, then it is (d, e)-continuous.

Proof Let $a \in X$, and let $\varepsilon > 0$ be given. We must find a $\delta > 0$ such that

$$e(f(x), f(a)) < \varepsilon \quad \text{whenever} \quad d(x, a) < \delta.$$

We first find an upper bound for $e(f(x), f(a))$ in terms of $d(x, a)$.

Since f is Lipschitz, there is a non-negative real number M such that, for all $x \in X$,

$$e(f(x), f(a)) \leq M d(x, a).$$

Thus, if $d(x, a) < \delta$, then

$$e(f(x), f(a)) < M\delta.$$

Now take $\delta = \varepsilon/(M + 1)$. Then, $e(f(x), f(a)) < \dfrac{M}{M+1}\varepsilon < \varepsilon$ whenever $d(x, a) < \delta$.

We use $M + 1$ to avoid problems when $M = 0$.

We conclude that f is continuous at a. Since a is an arbitrary point of X, f is continuous on X. ■

Thus all Lipschitz functions are continuous. But not all continuous functions are Lipschitz. For example, for the Euclidean metric on $\mathbb{R}$, the function $f: \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is continuous, but not Lipschitz. For if f were Lipschitz, then there would be a constant $M \geq 0$ such that, for all $a, b \in \mathbb{R}$,

We showed that f is continuous in Worked problem 2.1 of *Unit 1*.

$$|b^2 - a^2| \leq M|b - a|,$$

and so $|b + a| \leq M$ for $a \neq b$. In particular if $b = 0$ then we would have $|a| \leq M$ for all $a \in \mathbb{R}$, which is clearly false. Hence f is not Lipschitz.

$|b^2 - a^2| = |b + a|\,|b - a|$.

This means that the Lipschitz condition is more restrictive than continuity, but in fact many commonly occurring functions are Lipschitz. Moreover, when a function *is* Lipschitz, the easiest way to prove that it is continuous is often to prove that it is Lipschitz.

Worked problem 2.4

Show that $I: C[0, 1] \to \mathbb{R}$ defined by $I(f) = \int_0^1 f(x)\,dx$ is $(d_{\max}, d^{(1)})$-continuous.

Basic properties of integrals are summarized in the Handbook.

Solution

We show that I is a $(d_{\max}, d^{(1)})$-Lipschitz function, and deduce that it is continuous.

Suppose that $f, g \in C[0, 1]$. We need to find an upper bound for $d^{(1)}(I(f), I(g))$ in terms of $d_{\max}(f, g)$. Now

$$d^{(1)}(I(f), I(g)) = \left|\int_0^1 g(x)\,dx - \int_0^1 f(x)\,dx\right| = \left|\int_0^1 (g(x) - f(x))\,dx\right|. \quad (2.4)$$

The definition of $d_{\max}$ implies that, for all $x \in [0,1]$,

$$|g(x) - f(x)| \leq d_{\max}(f,g),$$

which is equivalent to

$$-d_{\max}(f,g) \leq g(x) - f(x) \leq d_{\max}(f,g).$$

Integrating the terms in this double inequality between 0 and 1, we obtain

$$\int_0^1 -d_{\max}(f,g)\,dx \leq \int_0^1 (g(x) - f(x))\,dx \leq \int_0^1 d_{\max}(f,g)\,dx;$$

that is,

$$-d_{\max}(f,g) \leq \int_0^1 (g(x) - f(x))\,dx \leq d_{\max}(f,g),$$

which is equivalent to

$$\left|\int_0^1 (g(x) - f(x))\,dx\right| \leq d_{\max}(f,g).$$

Therefore, by (2.4),

$$d^{(1)}(I(f), I(g)) \leq d_{\max}(f,g).$$

Hence I is $(d_{\max}, d^{(1)})$-Lipschitz (with $M = 1$), and so by Theorem 2.5 is continuous on $C[0,1]$. ■

Problem 2.7

Show that $F\colon C[0,1] \to \mathbb{R}$ defined by $F(f) = f(0)$ is $(d_{\max}, d^{(1)})$-continuous.

3 New metrics from old

After working through this section, you should be able to:

- use a one–one map to transfer a metric from one set W to another set X;
- given a metric on a set X, describe the *induced metric* on a subset of X;
- construct a metric on the product of metric spaces;
- recognize that different metrics can give rise to the same continuous functions.

In the previous section, we saw some examples of metrics and metric spaces. In this section, we show how we can use a given metric space to create a new metric space, on a different set or on a subset of the underlying set of the given metric space. We also show how to combine metric spaces to form new ones. These methods for creating new metric spaces comprise Subsections 3.1 and 3.2. In Subsection 3.3 you will see that the ways we suggest for combining metric spaces do not necessarily lead to different sets of continuous functions.

3.1 Transferring a metric

Our first method for creating new metric spaces is to use functions to transfer metrics from one space to another.

Suppose that we have a set X on which we wish to define a metric, and that we also have a metric space (W, d). We wish to 'transfer' the metric d to X; that is, we wish to define a metric on X in terms of the metric d on W. At first sight, it may seem that what we need to do is to bring the metric d from W to X by defining a function from W to X. Rather surprisingly, this does not work effectively; but a suitable function *from X to W* does the trick.

So, suppose that $f: X \to W$ and consider two points, x and y, in X. Since f maps from X to W, there are points v and w in W for which $v = f(x)$ and $w = f(y)$. We could define $d_f(x, y) = d(v, w) = d(f(x), f(y))$ and hope that this gives us a metric. As we shall see, for this to work, we require $f: X \to W$ to be a *one–one* function. (Recall that f is *one–one* if $x = y$ whenever $f(x) = f(y)$.)

We first give the function d_f a name.

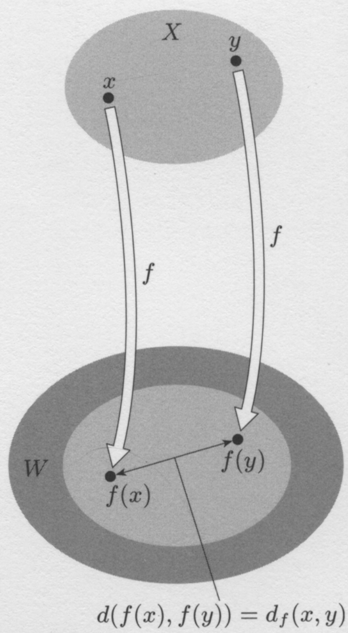


Figure 3.1

> ***Definition***
>
> Let $f: X \to W$ be a function and let d be a metric on W. The **pull-back of d by f** is the function $d_f: X \times X \to \mathbb{R}$ given by
>
> $$d_f(x, y) = d(f(x), f(y)), \quad \text{for all } x, y \in X.$$

We now prove that d_f is a metric when f is one–one.

> **Theorem 3.1**
>
> Let $f: X \to W$ be a one–one function and let d be a metric on W. Then the pull-back of d by f is a metric on X.

Proof We show that d_f satisfies conditions (M1)–(M3).

(M1) The metric d takes only non-negative values. So, for all $x, y \in X$, $d_f(x, y) = d(f(x), f(y)) \geq 0$.

For any $x \in X$, $d_f(x, x) = d(f(x), f(x)) = 0$.

Conversely, if $x, y \in X$ are such that $d_f(x, y) = 0$, then $d(f(x), f(y)) = 0$. Since d is a metric, this implies that $f(x) = f(y)$. But f is one–one, so $x = y$, and d_f satisfies (M1).

This is why f must be one–one.

(M2) For all $x, y \in X$,

$$d_f(y, x) = d(f(y), f(x)) = d(f(x), f(y)) = d_f(x, y),$$

since d is a metric on W. Thus, d_f satisfies (M2).

(M3) Let $x, y, z \in X$. Then, using property (M3) of d,

$$\begin{aligned} d_f(x, z) &= d(f(x), f(z)) \\ &\leq d(f(x), f(y)) + d(f(y), f(z)) \\ &= d_f(x, y) + d_f(y, z). \end{aligned}$$

Thus, d_f satisfies (M3).

Therefore d_f satisfies (M1)–(M3), and so is a metric on X. ■

Remark

If f is not one–one, then there are $x, y \in X$ such that $x \neq y$ and $f(x) = f(y)$. For this choice of x and y, we have

$$d_f(x, y) = d(f(x), f(y)) = 0.$$

But this means that d_f fails to satisfy (M1), and so is not a metric. Thus the pull-back of d by f is a metric *if and only if* f is one–one.

Worked problem 3.1

Show that $d: (0, 1) \times (0, 1) \to \mathbb{R}$ given by $d(x, y) = |y^{-1} - x^{-1}|$ is a metric on the open interval $(0, 1)$.

Solution

We could show that d is a metric by verifying directly that it satisfies (M1)–(M3). However, it is easier to observe that, if $f: (0, 1) \to \mathbb{R}$ is given by $f(x) = x^{-1}$, then d is the pull-back of the Euclidean metric $d^{(1)}$ on $\mathbb{R}$ by f: this is because

$$d(x, y) = |y^{-1} - x^{-1}| = d^{(1)}(f(x), f(y)).$$

Thus, by Theorem 3.1 we need only verify that f is one–one to conclude that d is a metric.

To show that f is one–one, suppose that $x, y \in (0, 1)$ and that $f(x) = f(y)$. Then $x^{-1} = y^{-1}$, so that $x = y$. Hence f is one–one, and so d is a metric on $(0, 1)$. ■

Problem 3.1

Use Theorem 3.1 to show that $d: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by

$$d(x, y) = |\tan^{-1} y - \tan^{-1} x|$$

is a metric on $\mathbb{R}$.

Hint Remember that $\tan^{-1}$ is the inverse function to tan, so that $\tan^{-1} x$ is the unique angle $\theta \in (-\pi/2, \pi/2)$ such that $\tan \theta = x$. Its graph is sketched in Figure 3.2.

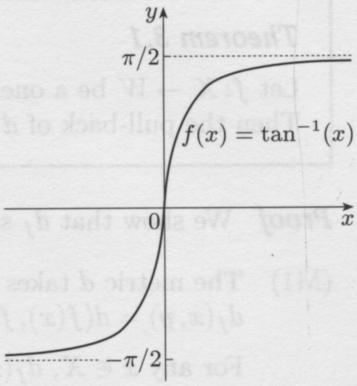


Figure 3.2

Remark

The metric defined in Problem 3.1 is referred to as the **$\tan^{-1}$ metric.**

There is an interesting geometric interpretation of the $\tan^{-1}$ metric. Consider Figure 3.3. A surveyor at the point P, one unit up the vertical axis, sights the points X and Y (both on the horizontal axis) along the lines PX and PY shown. With P as the vertex, these lines determine an angle ϕ. Let us use ϕ as a measure of the distance d (from the surveyor's perspective) between X and Y.

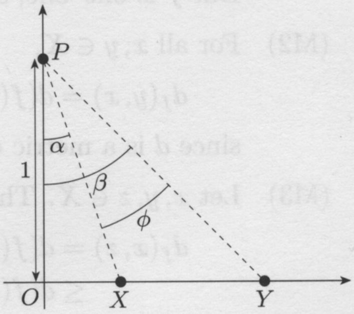


Figure 3.3

Let X and Y be at distances x and y respectively from the origin O. We now find d as a function of x and y. By considering the triangle with vertices P, O and X, we find $\tan \alpha = x$ (since $OP = 1$), so $\alpha = \tan^{-1} x$. Similarly, $\beta = \tan^{-1} y$. As illustrated, $y \geq x$ and $\phi = \beta - \alpha$, so $d(X, Y) = \phi = \tan^{-1} y - \tan^{-1} x$. In the case $x \geq y$, we would find that $d(X, Y) = \phi = \tan^{-1} x - \tan^{-1} y$. Combining these two cases yields the formula for $d(x, y)$:

$$d(X, Y) = |\tan^{-1} y - \tan^{-1} x|.$$

Thus, $d(X, Y)$ is the angle between the points X and Y, as observed by the surveyor.

Problem 3.2

Let $d: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by

$$d(x, y) = |y^2 - x^2|.$$

Determine whether d is a metric on $\mathbb{R}$.

Metric subspaces

Sometimes we wish to consider subsets A of the underlying set X of a metric space (X, d). Since $d(a, b)$ is defined for all points $a, b \in A$, you can easily check that we obtain a new metric space if we restrict d to the domain $A \times A$. The metric d restricted to this domain is denoted by d_A. On the one hand, (A, d_A) is a metric space in its own right. On the other hand, (A, d_A) is constructed from (X, d) and it is often useful to bear this in mind, in which case we refer to (A, d_A) as a *subspace* of (X, d).

Definition

Let (X, d) be a metric space and let $A \subseteq X$.

The function $d_A: A \times A \to \mathbb{R}$ given by

$$d_A(a, b) = d(a, b), \quad \text{for all } a, b \in A,$$

is the metric on A **induced by** d, often referred to as the **induced metric** on A.

The metric space (A, d_A) is a **metric subspace** of (X, d), often referred to as the **induced subspace**.

There is an interesting connection here with the pull-back construction. The function $f: A \to X$ given by

$$f(a) = a$$

is clearly one–one, and the pull-back metric on A obtained from f is

$$d_f(a, b) = d(f(a), f(b)) = d(a, b) = d_A(a, b) \quad \text{for all } a, b \in A.$$

That is, d_A is the same as the pull-back of d to A by the one–one function f.

The idea of an induced metric gives us a way to define a metric on common geometric surfaces in $\mathbb{R}^3$, such as the unit sphere with centre $\mathbf{0}$:

$$S = S_{d^{(3)}}(\mathbf{0}, 1) = \{\mathbf{x} \in \mathbb{R}^3 : d^{(3)}(\mathbf{0}, \mathbf{x}) = 1\}.$$

We simply endow S with the metric induced by the Euclidean metric $d^{(3)}$.

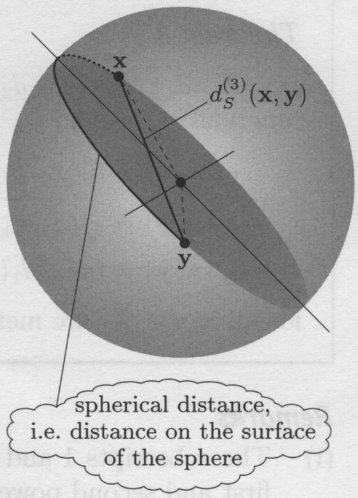


Figure 3.4

Problem 3.3

Show that $(1, 0, 0)$ and $(0, 0, 1)$ are in S, and calculate $d_S^{(3)}((1, 0, 0), (0, 0, 1))$.

Although endowing S with the metric induced by $d^{(3)}$ gives us a metric on the unit sphere, it does not measure distance *on* the sphere in the way that we do on the Earth, as Figure 3.4 illustrates. It is possible to define a metric that does this, but it is a little messy and we omit it.

Problem 3.4

A piece of string S, 11 cm long, is laid out so that it nearly encloses a square area of side 3 cm, leaving a gap in the middle of one side. Let d be the metric on S defined by distance in centimetres *along the string* and let $d_S^{(2)}$ be the metric on S induced by the Euclidean metric on $\mathbb{R}^2$, using centimetres as the units along both axes.

(a) Let a and b be the ends of the string. Find $d(a, b)$ and $d_S^{(2)}(a, b)$.

(b) What is the maximum possible value of $d_S^{(2)}(x, y)$ for any two points $x, y \in S$?

You may assume that d defines a metric. It is equivalent to the metric induced on S by the Euclidean metric $d^{(1)}$ when S is laid out along the real line, using centimetres as the units.

3.2 Product metrics

Earlier in this section we saw how to transfer a metric from one set to another (or to a subset of itself) by a straightforward one–one function. Suppose, however, that we have a set X defined as the *product* $X_1 \times X_2$ of two sets, each with a metric defined on it. Can we construct a metric on the product set X from the metrics on X_1 and X_2?

You have already come across two examples where this can be done, in both of which $X = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ and $X_1 = X_2 = \mathbb{R}$ with the Euclidean metric $d^{(1)}$. The first is the Euclidean metric $d^{(2)}$ on $\mathbb{R}^2$, which can be written in terms of the Euclidean metric $d^{(1)}$ on $\mathbb{R}$ as

$$\begin{aligned} d^{(2)}(\mathbf{x}, \mathbf{y}) &= \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} \\ &= \sqrt{(d^{(1)}(x_1, y_1))^2 + (d^{(1)}(x_2, y_2))^2}. \end{aligned}$$

The second is the taxicab metric on $\mathbb{R}^2$, which can be written in terms of the Euclidean metric $d^{(1)}$ on $\mathbb{R}$ as

$$e_1(\mathbf{x}, \mathbf{y}) = d^{(1)}(x_1, y_1) + d^{(1)}(x_2, y_2).$$

In fact, if (X_1, d_1) and (X_2, d_2) are metric spaces, then there are many metrics that we can construct on the product set $X = X_1 \times X_2$ from the metrics d_1 and d_2. Such metrics are referred to as **product metrics**. We look in detail at three of them.

Theorem 3.2

Let (X_1, d_1) and (X_2, d_2) be metric spaces, and let $X = X_1 \times X_2$. The functions e_1, e_2 and e_∞ from $X \times X$ to $\mathbb{R}$, given by

$$\begin{aligned} e_1(\mathbf{x}, \mathbf{y}) &= d_1(x_1, y_1) + d_2(x_2, y_2), \\ e_2(\mathbf{x}, \mathbf{y}) &= \sqrt{(d_1(x_1, y_1))^2 + (d_2(x_2, y_2))^2}, \\ e_\infty(\mathbf{x}, \mathbf{y}) &= \max\{d_1(x_1, y_1), d_2(x_2, y_2)\}, \end{aligned}$$

for all $\mathbf{x}, \mathbf{y} \in X$, are metrics on X.

Remarks

(i) The subscripts 1 and 2 indicate that e_1 and e_2 are formed by using first and second powers. More generally, if we consider an arbitrary power $p \geq 1$ and define

$$e_p(\mathbf{x}, \mathbf{y}) = (d_1(x_1, y_1)^p + d_2(x_2, y_2)^p)^{1/p},$$

then e_p is also a metric on X. Moreover, if we take the limit as $p \to \infty$, then we find that $e_p(\mathbf{x}, \mathbf{y}) \to e_\infty(\mathbf{x}, \mathbf{y})$.

(ii) In the proof that e_∞ is a metric, we use the following fact:

if $a, b, c, d \in \mathbb{R}$, then $\max\{a + b, c + d\} \leq \max\{a, c\} + \max\{b, d\}$.

This follows easily by considering $a + b$ and $c + d$ separately, and by noting that each of these is at most $\max\{a, c\} + \max\{b, d\}$.

Proof The proof that e_1 is a metric is very similar to the proof (given in Subsection 1.1) that the taxicab metric on $\mathbb{R}^2$ is a metric.

You are asked to provide this proof in Problem 3.5.

The proof that e_2 is a metric is very similar to the proof (given in *Unit A1*) that $d^{(2)}$ satisfies conditions (M1)–(M3).

You are asked to provide this proof in the problems for this unit.

Here we verify that e_∞ satisfies conditions (M1)–(M3).

(M1) For all $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$, $d_1(x_1, y_1)$ and $d_2(x_2, y_2)$ are non-negative, since d_1 and d_2 are metrics, and hence so is their maximum: thus $e_\infty(\mathbf{x}, \mathbf{y}) \geq 0$.

For all $\mathbf{x} \in X_1 \times X_2$,

$$e_\infty(\mathbf{x}, \mathbf{x}) = \max\{d_1(x_1, x_1),\ d_2(x_2, x_2)\} = \max\{0, 0\} = 0,$$

since d_1 and d_2 are metrics.

Conversely, suppose $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$ are such that $e_\infty(\mathbf{x}, \mathbf{y}) = 0$. Then

$$0 = \max\{d_1(x_1, y_1),\ d_2(x_2, y_2)\}.$$

Since d_1 and d_2 are metrics, $d_1(x_1, y_2) \geq 0$ and $d_2(x_1, y_2) \geq 0$, so this equation implies that $d_1(x_1, y_1) = 0$ and $d_2(x_2, y_2) = 0$. Thus $x_1 = y_1$ and $x_2 = y_2$, since d_1 and d_2 are metrics; that is, $\mathbf{x} = \mathbf{y}$.

Therefore e_∞ satisfies (M1).

(M2) Let $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$. Using property (M2) for the metrics d_1 and d_2, we have

$$\begin{aligned} e_\infty(\mathbf{y}, \mathbf{x}) &= \max\{d_1(y_1, x_1),\ d_2(y_2, x_2)\} \\ &= \max\{d_1(x_1, y_1),\ d_2(x_2, y_2)\} = e_\infty(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Thus e_∞ satisfies (M2).

(M3) Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X_1 \times X_2$. Then, using property (M3) for the metrics d_1 and d_2,

$$\begin{aligned} e_\infty(\mathbf{x}, \mathbf{z}) &= \max\{d_1(x_1, z_1),\ d_2(x_2, z_2)\} \\ &\leq \max\{d_1(x_1, y_1) + d_1(y_1, z_1),\ d_2(x_2, y_2) + d_2(y_2, z_2)\} \\ &\leq \max\{d_1(x_1, y_1),\ d_2(x_2, y_2)\} + \max\{d_1(y_1, z_1),\ d_2(y_2, z_2)\}. \\ &= e_\infty(\mathbf{x}, \mathbf{y}) + e_\infty(\mathbf{y}, \mathbf{z}). \end{aligned}$$

This step is justified by Remark (ii) preceding this proof.

Thus e_∞ satisfies (M3).

Therefore e_∞ is a metric on $X_1 \times X_2$. ■

Problem 3.5

Let (X_1, d_1) and (X_2, d_2) be metric spaces. Prove that $(X_1 \times X_2, e_1)$ is a metric space.

Product metrics on the plane

When $(X_1, d_1) = (X_2, d_2) = (\mathbb{R}, d^{(1)})$, our methods for constructing metrics on product spaces give us three different ways of defining a metric on $\mathbb{R}^2$:

$$\begin{aligned} e_1(\mathbf{x}, \mathbf{y}) &= d^{(1)}(x_1, y_1) + d^{(1)}(x_2, y_2) = |y_1 - x_1| + |y_2 - x_2|; \\ e_2(\mathbf{x}, \mathbf{y}) &= \sqrt{d^{(1)}(x_1, y_1)^2 + d^{(1)}(x_2, y_2)^2} = \sqrt{|y_1 - x_1|^2 + |y_2 - x_2|^2}; \\ e_\infty(\mathbf{x}, \mathbf{y}) &= \max\{d^{(1)}(x_1, y_1), d^{(1)}(x_2, y_2)\} = \max\{|y_1 - x_1|, |y_2 - x_2|\}. \end{aligned}$$

The metrics e_1 and e_2 are the familiar taxicab and Euclidean metrics on the plane. The metric e_∞ is known as the **max metric** on the plane.

Problem 3.6

Show that e_1, e_2 and e_∞ are distinct metrics on the plane, by calculating $e_1(\mathbf{0},(2,1))$, $e_2(\mathbf{0},(2,1))$ and $e_\infty(\mathbf{0},(2,1))$.

We can obtain an idea of the difference in behaviour between these metrics on the plane by sketching their unit open balls centred at the origin, as shown in Figure 3.5. In Worked problem 1.1, we saw that for e_1, the taxicab metric, the unit open ball centred at the origin consists of all the points inside a diamond centred at the origin. In Subsection 1.2, we saw that for e_2, the Euclidean metric $d^{(2)}$, the unit open ball centred at the origin is an open disc of unit radius centred at the origin. For e_∞, the unit open ball centred at the origin consists of all the points inside the square of side length 2 centred at the origin. In all three cases, since we are considering *open* balls, the boundary points are excluded.

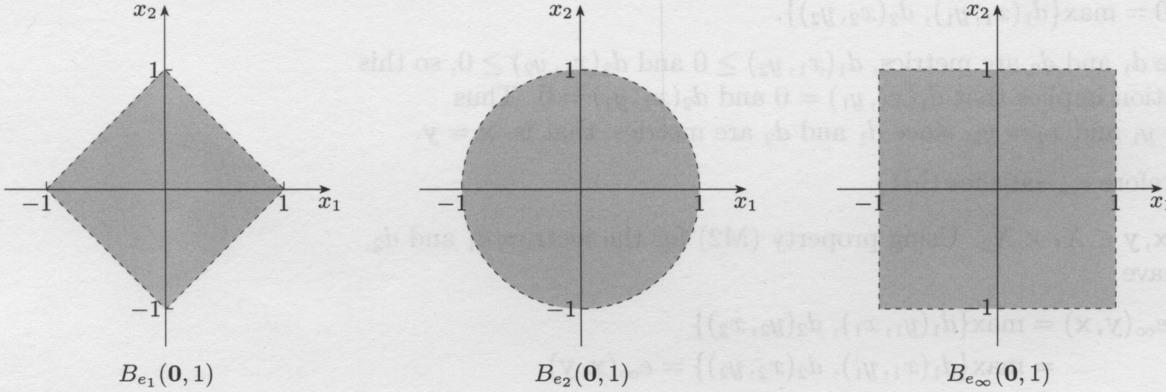


Figure 3.5

Problem 3.7

(a) Justify the sketch of the unit open ball centred at the origin for e_∞.

(b) What are the unit closed balls and spheres centred at the origin for e_1, e_2 and e_∞?

3.3 Product metrics and continuity

Consider the functions $f:\mathbb{R}\to\mathbb{R}$ and $g:\mathbb{R}\to\mathbb{R}$ given by $f(x)=x^2$ and $g(x)=x^3$; both f and g are $(d^{(1)},d^{(1)})$-continuous. Now consider the function $h:\mathbb{R}^2\to\mathbb{R}^2$ defined by

$$h((x_1,x_2))=(f(x_1),g(x_2))=(x_1^2,x_2^3).$$

Is this function continuous?

This question makes no sense unless we define a metric on $\mathbb{R}^2$. It seems reasonable to use a product metric of $d^{(1)}$ with $d^{(1)}$, but we have a choice of three. Which one should we choose? It turns out that it does not matter which metric we choose — they all result in the same set of continuous functions.

One way to understand why this is so is to look carefully at the unit open balls for the metrics e_1, e_2 and e_∞. You should be able to see that $B_{e_1}(\mathbf{0},1)$ fits inside $B_{e_2}(\mathbf{0},1)$, which fits inside $B_{e_\infty}(\mathbf{0},1)$. Moreover, $B_{e_\infty}(\mathbf{0},1)$ fits inside a scaled version of $B_{e_1}(\mathbf{0},1)$, the ball $B_{e_1}(\mathbf{0},2)$. This is illustrated in Figure 3.6. These inclusions can be generalized to open balls

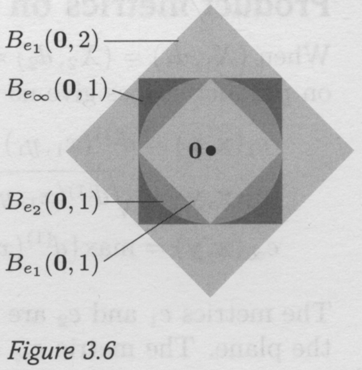


Figure 3.6

with any given centre a and radius r. Thus, as you may be able to deduce, if we can show continuity using open balls for one of the metrics, then we can find suitable open balls to show continuity for the other two.

To prove this formally, we first need to establish some inequalities between the three metrics.

Worked problem 3.2

Show that, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$:

(a) $e_2(\mathbf{x}, \mathbf{y}) \le e_1(\mathbf{x}, \mathbf{y})$;

(b) $e_\infty(\mathbf{x}, \mathbf{y}) \le e_2(\mathbf{x}, \mathbf{y})$;

(c) $e_1(\mathbf{x}, \mathbf{y}) \le 2e_\infty(\mathbf{x}, \mathbf{y})$.

Solution

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$.

(a) We use the fact that, if a and b are non-negative real numbers, then $a^2 + b^2 \le a^2 + 2ab + b^2 = (a + b)^2$, and so $\sqrt{a^2 + b^2} \le a + b$.

$$\begin{aligned} e_2(\mathbf{x}, \mathbf{y}) &= \sqrt{|y_1 - x_1|^2 + |y_2 - x_2|^2} \\ &\le |y_1 - x_1| + |y_2 - x_2| = e_1(\mathbf{x}, \mathbf{y}). \end{aligned}$$

(b) $$\begin{aligned} e_\infty(\mathbf{x}, \mathbf{y}) &= \max\{|y_1 - x_1|, |y_2 - x_2|\} \\ &\le \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} = e_2(\mathbf{x}, \mathbf{y}). \end{aligned}$$

(c) $$\begin{aligned} e_1(\mathbf{x}, \mathbf{y}) &= |y_1 - x_1| + |y_2 - x_2| \\ &\le \max\{|y_1 - x_1|, |y_2 - x_2|\} + \max\{|y_1 - x_1|, |y_2 - x_2|\} \\ &= 2\max\{|y_1 - x_1|, |y_2 - x_2|\} = 2e_\infty(\mathbf{x}, \mathbf{y}). \end{aligned}$$ ■

These inequalities imply that, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$,

$$e_2(\mathbf{x}, \mathbf{y}) \le e_1(\mathbf{x}, \mathbf{y}) \le 2e_2(\mathbf{x}, \mathbf{y}), \tag{3.1}$$

$$e_\infty(\mathbf{x}, \mathbf{y}) \le e_2(\mathbf{x}, \mathbf{y}) \le 2e_\infty(\mathbf{x}, \mathbf{y}), \tag{3.2}$$

$$\tfrac{1}{2}e_1(\mathbf{x}, \mathbf{y}) \le e_\infty(\mathbf{x}, \mathbf{y}) \le e_1(\mathbf{x}, \mathbf{y}). \tag{3.3}$$

Metrics that are related in this way are said to be *metrically equivalent.*

Definition

Let d_1 and d_2 be two metrics defined on a set X. The metrics d_1 and d_2 are **metrically equivalent** if there are positive real numbers m and M such that, for all $x, y \in X$,

$$md_1(x, y) \le d_2(x, y) \le Md_1(x, y). \tag{3.4}$$

In *Unit A3* you will meet a different concept of equivalence between metrics.

Remarks

(i) Notice that inequalities (3.4) are equivalent to

$$\tfrac{1}{M}d_2(x, y) \le d_1(x, y) \le \tfrac{1}{m}d_2(x, y)$$

and so there is no asymmetry in the definition: saying that d_1 is metrically equivalent to d_2 is the same as saying that d_2 is metrically equivalent to d_1.

(ii) Inequalities (3.1)–(3.3) indicate that the metrics e_1, e_2 and e_∞ on $\mathbb{R}^2$ are metrically equivalent.

The argument used to obtain inequalities (3.1)–(3.3) works not only for the product metrics on $\mathbb{R}^2$ but also for the metrics e_1, e_2 and e_∞ defined

for the general product $X_1 \times X_2$. We omit the details, which are not very different from those for $\mathbb{R}^2$, and simply record the result.

Theorem 3.3

Let (X_1, d_1) and (X_2, d_2) be metric spaces, and let $X = X_1 \times X_2$. Then the three product metrics e_1, e_2 and e_∞ on X are metrically equivalent.

Problem 3.8

Show that e_1 is metrically equivalent on $\mathbb{R}^2$ to the metric e given by

$$e(\mathbf{x}, \mathbf{y}) = |y_1 - x_1| + 2|y_2 - x_2|.$$

You may assume that e is a metric on the plane.

The usefulness of metrically equivalent metrics lies in the following result, which tells us that it does not matter which of them we use in testing a function for continuity.

Theorem 3.4

Let (Y, e) be a metric space and let d_1 and d_2 be metrically equivalent metrics on a set X.

If $f: X \to Y$ and $a \in X$, then

f is (d_1, e)-continuous at a if and only if f is (d_2, e)-continuous at a.

If $g: Y \to X$ and $b \in Y$, then

g is (e, d_1)-continuous at b if and only if g is (e, d_2)-continuous at b.

Proof Since d_1 and d_2 are metrically equivalent metrics on X, there are positive real numbers m and M such that, for all $u, v \in X$,

$$m d_1(u, v) \le d_2(u, v) \le M d_1(u, v). \tag{3.5}$$

We prove that, if $f: X \to Y$ and $a \in X$, then

f is (d_1, e)-continuous at a if and only if f is (d_2, e)-continuous at a.

The proof for $g: Y \to X$ is similar.

You are asked to prove this in the problems for this unit.

Proof that if f is (d_1, e)-continuous at a then f is (d_2, e)-continuous at a

Suppose that f is (d_1, e)-continuous at the point $a \in X$ and let $\varepsilon > 0$ be given. Since f is (d_1, e)-continuous at a, there is a $\delta_1 > 0$ such that

$$e(f(a), f(x)) < \varepsilon \quad \text{whenever} \quad d_1(a, x) < \delta_1.$$

In order to show that f is (d_2, e)-continuous at a, we must find a $\delta_2 > 0$ such that $e(f(a), f(x)) < \varepsilon$ whenever $d_2(a, x) < \delta_2$. We find δ_2 by seeking an upper bound for $d_1(a, x)$ in terms of $d_2(a, x)$.

Such an upper bound follows immediately from the left-hand inequality of (3.5), for this implies that

$$d_1(a, x) \le \tfrac{1}{m} d_2(a, x).$$

We therefore set $\delta_2 = m\delta_1$. If $d_2(a, x) < \delta_2$, then

$$d_1(a, x) \le \tfrac{1}{m} d_2(a, x) < \tfrac{1}{m}\delta_2 = \delta_1,$$

and so $e(f(a), f(x)) < \varepsilon$.

Thus, f is (d_2, e)-continuous at a, as required.

Proof that if f is (d_2, e)-continuous at a then f is (d_1, e)-continuous at a

Suppose that f is (d_2, e)-continuous at the point $a \in X$ and let $\varepsilon > 0$ be given. Since f is (d_2, e)-continuous at a, there is a $\delta_2 > 0$ such that

$$e(f(a), f(x)) < \varepsilon \quad \text{whenever} \quad d_2(a, x) < \delta_2.$$

In order to show that f is (d_1, e)-continuous at a, we must find a $\delta_1 > 0$ such that $e(f(a), f(x)) < \varepsilon$ whenever $d_1(a, x) < \delta_1$. We find δ_1 by seeking an upper bound for $d_2(a, x)$ in terms of $d_1(a, x)$.

Such an upper bound follows immediately from the right-hand inequality of (3.5), for this implies that

$$d_2(a, x) \leq M d_1(a, x).$$

We therefore set $\delta_1 = \frac{1}{M}\delta_2$. If $d_1(a, x) < \delta_1$, then

$$d_2(a, x) \leq M d_1(a, x) < M\delta_1 = \delta_2,$$

and so $e(f(a), f(x)) < \varepsilon$.

Thus, f is (d_1, e)-continuous at a, as required. ■

Since, by Theorem 3.3, the metrics e_1, e_2 and e_∞ are metrically equivalent on any product $X_1 \times X_2$, the following result follows directly from Theorem 3.4.

Theorem 3.5

Let (X_1, d_1), (X_2, d_2) and (Y, e) be metric spaces, let $X = X_1 \times X_2$ and let e_1, e_2 and e_∞ be the three product metrics on X.

If $f: X \to Y$, then, for $j, k = 1, 2, \infty$,

f is (e_j, e)-continuous if and only if f is (e_k, e)-continuous.

If $g: Y \to X$, then, for $j, k = 1, 2, \infty$,

g is (e, e_j)-continuous if and only if g is (e, e_k)-continuous.

Remark

Suppose we have a function f from or to $X_1 \times X_2$ with one of the product metrics e_1, e_2 or e_∞. Theorem 3.5 tells us that we can test f for continuity by using any one of e_1, e_2 or e_∞, and so it allows us to make the simplest choice.

Projection functions and continuity

In Section 5 of *Unit A1*, we saw that we can show that a function $f: \mathbb{R}^2 \to \mathbb{R}^2$ is continuous by showing that both $p_1 \circ f: \mathbb{R}^2 \to \mathbb{R}$ and $p_2 \circ f: \mathbb{R}^2 \to \mathbb{R}$ are continuous, where $p_1: \mathbb{R}^2 \to \mathbb{R}$ and $p_2: \mathbb{R}^2 \to \mathbb{R}$ are the projection functions defined by $p_1(x_1, x_2) = x_1$ and $p_2(x_1, x_2) = x_2$. We can extend the definitions of projection functions to a general product space, and a similar result about continuity holds.

Definition

Let X_1 and X_2 be sets and let $X = X_1 \times X_2$. The **projection functions** are the functions $p_1: X \to X_1$ and $p_2: X \to X_2$ given by

$$p_1(x_1, x_2) = x_1, \qquad p_2(x_1, x_2) = x_2.$$

As with the projection functions in *Unit A1*, these functions are continuous.

Theorem 3.6

Let (X_1, d_1) and (X_2, d_2) be metric spaces and let (X, e) be the metric space where $X = X_1 \times X_2$ and e is one of the product metrics e_1, e_2 or e_∞. Then p_1 is (e, d_1)-continuous and p_2 is (e, d_2)-continuous, where p_1 and p_2 are projection functions from X to X_1 and X_2 respectively.

Proof Once we have proved the result for any one of e_1, e_2 or e_∞, the others follow immediately from Theorem 3.5. We prove the result for e_∞, as this turns out to be the simplest to work with. We show that p_1 is (e_∞, d_1)-continuous; the proof that p_2 is (e_∞, d_2)-continuous is similar.

Choose $\mathbf{a} \in X$. We find an upper bound on $d_1(p_1(\mathbf{a}), p_1(\mathbf{x}))$ in terms of $e_\infty(\mathbf{a}, \mathbf{x})$:

$$\begin{aligned} d_1(p_1(\mathbf{a}), p_1(\mathbf{x})) &= d_1(a_1, x_1) \\ &\le \max\{d_1(a_1, x_1), d_2(a_2, x_2)\} = e_\infty(\mathbf{a}, \mathbf{x}). \end{aligned}$$

Thus p_1 is an (e_∞, d_1)-Lipschitz function, and so, by Theorem 2.5, p_1 is (e_∞, d_1)-continuous on X. ■

We can now derive the promised continuity result. Since its proof is similar to that of Theorem 5.5 of *Unit A1*, we omit it.

Theorem 3.7

Let (Y, d), (X_1, d_1) and (X_2, d_2) be metric spaces, and let (X, e) be the metric space where $X = X_1 \times X_2$ and e is one of the product metrics e_1, e_2 or e_∞. Then a function $f: Y \to X_1 \times X_2$ is (d, e)-continuous at $a \in Y$ if and only if $p_1 \circ f: Y \to X_1$ is (d, d_1)-continuous at a and $p_2 \circ f: Y \to X_2$ is (d, d_2)-continuous at a.

Worked problem 3.3

Consider the metric spaces $(C[0,1], d_{\max})$ and $(\mathbb{R}, d^{(1)})$. Define the function $F: C[0,1] \to \mathbb{R} \times C[0,1]$ by

$$F(f) = (f(0), f).$$

Show that F is $(d_{\max}, e)$-continuous where e denotes any of e_1, e_2 or e_∞.

Solution

Consider the composites of the projection functions with F, that is, $p_1 \circ F: C[0,1] \to \mathbb{R}$ and $p_2 \circ F: C[0,1] \to C[0,1]$. Now

$$(p_1 \circ F)(f) = f(0),$$

and we saw in Problem 2.7 that this is a $(d_{\max}, d^{(1)})$-continuous function. Also,

$$(p_2 \circ F)(f) = f,$$

and so

$$d_{\max}((p_2 \circ F)(f), (p_2 \circ F)(g)) = d_{\max}(f, g).$$

Thus, $p_2 \circ F$ is a $(d_{\max}, d_{\max})$-Lipschitz function (with $M = 1$), and so by Theorem 2.5 it is $(d_{\max}, d_{\max})$-continuous.

Since both $p_1 \circ F$ and $p_2 \circ F$ are continuous, Theorem 3.7 implies that F is $(d_{\max}, e)$-continuous, where e is any of e_1, e_2 or e_∞. ■

Problem 3.9

Consider the metric spaces $(\mathbf{C}, d_{\mathbf{C}})$ and $(\mathbb{R}, d^{(1)})$. Define the function $f: \mathbf{C} \to \mathbb{R} \times \mathbf{C}$ by

$$f(\mathbf{x}) = (x_1, \sigma(\mathbf{x})).$$

(a) Evaluate $f(\mathbf{x})$ when:

(i) $\mathbf{x} = (0, 0, 0, 0, 0, 0, \ldots)$;

(ii) $\mathbf{x} = (0, 1, 0, 1, 0, 1, \ldots)$.

(b) Show that f is $(d_{\mathbf{C}}, e)$-continuous where e denotes any of e_1, e_2 or e_∞.

You may assume that $g: \mathbf{C} \to \mathbb{R}$ given by $g(\mathbf{x}) = x_1$ is $(d_{\mathbf{C}}, d^{(1)})$-continuous.

n-fold products

Suppose that we are interested in the product of three metric spaces. In principle, this is no more difficult than the product of two: we take the product of the first two, which is a metric space; then we take the product of this new metric space with the third. In doing this, we may use e_1, e_2 or e_∞, or any other product metric.

This construction can be continued, enabling us to obtain products of n metric spaces, for any $n \in \mathbb{N}$. In doing this, we find that there are appropriate generalizations of Theorems 3.5, 3.6 and 3.7.

4 Open sets

After working through this section, you should be able to:

- state and use the *Fried-egg Property* for balls;
- state the definition of an *open set*;
- determine whether a given set is open;
- outline the basic properties of open sets;
- define continuity in terms of open sets.

In the previous section, you saw that it is possible to change the metric on a set X without changing the continuity status of functions from and to X. You also saw that there are several different metrics defined on the product of two metric spaces, and that when these are metrically equivalent they result in the same continuous functions. This suggests that there may be something deeper underlying the notion of continuity that metrics do not capture. Indeed there is, and in this section we tease it out.

4.1 Introducing open sets

Our goal is to come to a deeper understanding of the notion of continuity. To do this our starting point will be the definition of continuity in terms of balls. Recall from Subsection 1.2 that, if (X,d) and (Y,e) are metric spaces, this definition states that a function $f\colon X \to Y$ is continuous at $a \in X$ if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$f(x) \in B_e(f(a),\varepsilon) \quad \text{whenever} \quad x \in B_d(a,\delta).$$

We now need to look more closely at the open balls $B_e(f(a),\varepsilon)$ and $B_d(a,\delta)$ in this definition. To do this, it is helpful to reformulate the definition in terms of the *inverse image* of a set.

Definition

Suppose that $f\colon X \to Y$ and $B \subseteq Y$. We define the **inverse image** of B under f to be the *set*

$$f^{-1}(B) = \{x \in X : f(x) \in B\}.$$

Remarks

(i) $f^{-1}(B)$ is a *set*, consisting of those points in X that are mapped by f into B.

(ii) If $A \subseteq X$ and $A \subseteq f^{-1}(B)$, then $f(x) \in B$ for all $x \in A$.

Problem 4.1

Let $f\colon \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$. Write down $f^{-1}(B)$ when:

(a) $B = \{0\}$; (b) $B = [0,4]$; (c) $B = [1,4)$; (d) $B = (-4,-1)$.

We now reformulate the definition of continuity in terms of open balls and inverse images.

> ***Definition***
>
> Let (X, d) and (Y, e) be metric spaces.
>
> A function $f \colon X \to Y$ is **continuous** at $a \in X$ if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that
>
> $$B_d(a, \delta) \subseteq f^{-1}(B_e(f(a), \varepsilon)).$$
>
> A function that is continuous at all points of X is **continuous** on X.

Remark (ii) above tells us that this definition is equivalent to the definition in terms of balls.

Remark

We shall refer to this definition of continuity on metric spaces as the **inverse image definition**.

It is instructive to examine more closely what this definition means, and a good place to start is inside an open ball. The 'openness' means that, *whatever point you choose in the ball, it must be at the centre of a smaller open ball that lies entirely inside the original ball* (as Figure 4.1 illustrates). We call this the *Fried-egg Property*, because you can imagine the outer ball as a fried egg and the inner ball as its (generally off-centre) yolk.

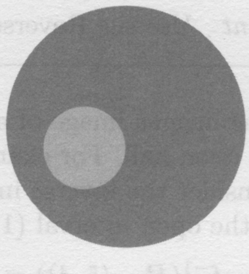


Figure 4.1

> ***Theorem 4.1 Fried-egg Property for balls***
>
> Let (X, d) be a metric space, let $a \in X$ and $r > 0$, and consider the open ball $B_d(a, r)$. Then, whenever $x \in B_d(a, r)$, we can find an $s > 0$ such that
>
> $$B_d(x, s) \subseteq B_d(a, r).$$

Example 4.1

Consider the open ball $B_{d^{(1)}}(2, 1)$ in $(\mathbb{R}, d^{(1)})$: that is, the open interval $(1, 3)$. Given any $x \in (1, 3)$, we can find $B_{d^{(1)}}(x, s)$, that is, the open interval $(x - s, x + s)$, such that

$$(x - s, x + s) \subseteq (1, 3).$$

For instance, if $x = 2.99$, then we may choose $s = 0.005$, since $(2.985, 2.995) \subseteq (1, 3)$. ■

We now prove Theorem 4.1.

Proof Let $x \in B_d(a, r)$, so $d(a, x) < r$. Now let $s = r - d(a, x) > 0$. Then, if $y \in X$ and $d(x, y) < s$, we may use property (M3) (the Triangle Inequality) for d to conclude that

$$d(a, y) \leq d(a, x) + d(x, y) < d(a, x) + s = r.$$

That is, $y \in B_d(a, r)$. Thus $B_d(x, s) \subseteq B_d(a, r)$. ■

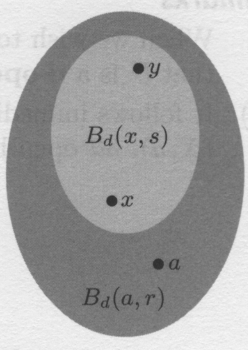


Figure 4.2

Remark

The closer to the boundary x is, the smaller the radius s has to be: the crucial point is that *there is always room enough.*

The Fried-egg Property tells us that we can enclose each point *inside* a given open ball within another (small enough) open ball lying entirely inside the original open ball. But we sometimes need to know that a point *outside* a given open ball can be enclosed in another (small enough) open ball lying entirely outside the original open ball. In the following problem we ask you to show that this can be done.

Problem 4.2

Let (X, d) be a metric space and let $r > 0$. Let a and x be points of X for which $d(a, x) > r$. Show that there is an $s > 0$ for which

$$B_d(a, r) \cap B_d(x, s) = \varnothing.$$

Hint Use the Reverse Triangle Inequality.

The inverse image of an open ball under a continuous function need not be an open ball. For example, let $f\colon \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$, and consider the inverse image of the open ball $B_{d^{(1)}}(5, 4)$. The open ball itself is the open interval $(1, 9)$, and so

$$f^{-1}(B_{d^{(1)}}(5, 4)) = f^{-1}((1, 9)) = (-3, -1) \cup (1, 3),$$

which is not an open ball in $(\mathbb{R}, d^{(1)})$. However, this set still possesses the Fried-egg Property: about any point in $(-3, -1) \cup (1, 3)$, we can find an open ball completely contained inside $(-3, -1) \cup (1, 3)$, as Figure 4.3 illustrates.

It is in fact the union of two open balls.

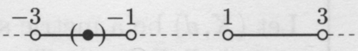


Figure 4.3

Sets that have the Fried-egg Property are important, so we give them a special name.

> ***Definition***
>
> Let (X, d) be a metric space and let A be a subset of X.
>
> A point $a \in A$ has the **Fried-egg Property** for d in A if there is an $r > 0$ such that
>
> $$B_d(a, r) \subseteq A.$$
>
> A subset U of X is an **open set** if each point $u \in U$ has the Fried-egg Property for d in U.

We often use U to denote an open set.

Remarks

(i) When we wish to emphasize the particular metric d on X, we say that U is a ***d*-open set** or that the set U is **open** in (X, d).

(ii) It follows immediately from Theorem 4.1 that, in a metric space (X, d), *all* open balls $B_d(a, r)$ are d-open sets.

Figure 4.4 shows three examples of open sets in the plane with the Euclidean metric.

The boundaries of the sets are drawn using broken lines to indicate that points on the boundaries are *not* included in the sets. We often use such broken lines to indicate the boundaries of open sets.

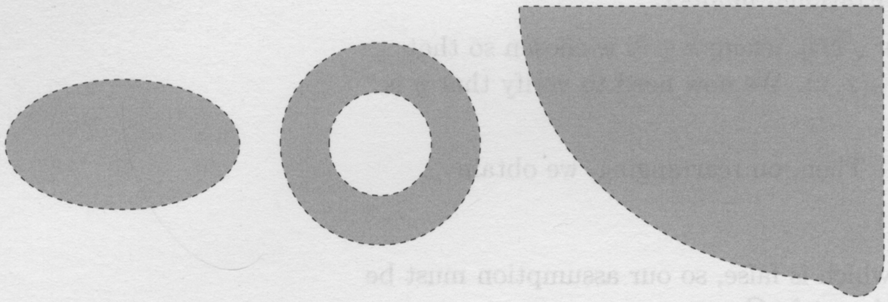


Figure 4.4

Open sets in Euclidean spaces

Consider $(\mathbb{R}, d^{(1)})$. If $a < b$, the open interval (a, b) in $\mathbb{R}$ is a $d^{(1)}$-open set: indeed, it is the open ball $B_{d^{(1)}}(\frac{1}{2}(a+b), \frac{1}{2}(b-a))$.

But what about intervals of the form $[a, b)$, $(a, b]$ or $[a, b]$?

Consider the interval $[a, b)$, where $a < b$. This set is not open, as the point a does not have the Fried-egg Property in $[a, b)$: for any $r > 0$, the ball $B_{d^{(1)}}(a, r) = (a - r, a + r)$ contains points $x < a$ (for example $a - \frac{1}{2}r$), and such points do not belong to $[a, b)$.

Problem 4.3

Explain why the sets $(a, b]$ and $[a, b]$ are not open.

Problem 4.4

Show that the intervals (a, ∞), $(-\infty, a)$ and $(-\infty, \infty) = \mathbb{R}$ are open sets, and explain why $(-\infty, a]$ and $[a, \infty)$ are not.

Now let us consider other subsets of $(\mathbb{R}, d^{(1)})$.

Any *finite* non-empty set of real numbers $A = \{a_1, a_2, \ldots, a_n\}$ is not open in $(\mathbb{R}, d^{(1)})$: in any open ball centred on a point in A, there are points not in A. Similarly, any infinite set of real numbers of the form

$$A = \{a_1, a_2, a_3, \ldots : a_1 < a_2 < a_3 < \cdots\}$$

cannot be open in $(\mathbb{R}, d^{(1)})$. To see this, consider any point a_k of A: for any $r > 0$, $B_{d^{(1)}}(a_k, r)$ contains points not in A. Thus, for example, the set $\mathbb{N} = \{1, 2, 3, \ldots\}$ of natural numbers is not open in $(\mathbb{R}, d^{(1)})$.

Worked problem 4.1

Show that the set $\mathbb{Q}$ of rational numbers is not open in $(\mathbb{R}, d^{(1)})$.

Solution

Let $x \in \mathbb{Q}$. We must show that, for any $r > 0$, there is a point $y \in B_{d^{(1)}}(x, r)$ such that $y \notin \mathbb{Q}$ — that is, each open ball centred on a rational number must contain an irrational number.

So, let $r > 0$ and consider $y = x + \sqrt{2}/k$, where $k \in \mathbb{N}$ is chosen so that $\sqrt{2}/k < r$. It follows that $y \in B_{d^{(1)}}(x, r)$. We now need to verify that y is irrational.

Suppose $y = x + \sqrt{2}/k$ is rational. Then, on rearranging, we obtain

$$k(y - x) = \sqrt{2}.$$

This implies that $\sqrt{2}$ is rational, which is false, so our assumption must be false. Thus, y is irrational and so is not in $\mathbb{Q}$.

Hence $B_{d^{(1)}}(x, r) \nsubseteq \mathbb{Q}$. Since x is an arbitrary point of $\mathbb{Q}$, it follows that $\mathbb{Q}$ is not open. ■

Let us now turn to subsets of the plane $\mathbb{R}^2$.

Worked problem 4.2

Consider the plane $\mathbb{R}^2$ with the Euclidean metric $d^{(2)}$.

(a) Show that

$$U = (0, 1) \times (0, 3) = \{\mathbf{x} \in \mathbb{R}^2 : 0 < x_1 < 1, 0 < x_2 < 3\}$$

is a $d^{(2)}$-open set.

(b) Show that

$$H = [0, 1] \times [0, 3] = \{\mathbf{x} \in \mathbb{R}^2 : 0 \le x_1 \le 1, 0 \le x_2 \le 3\}$$

is not a $d^{(2)}$-open set.

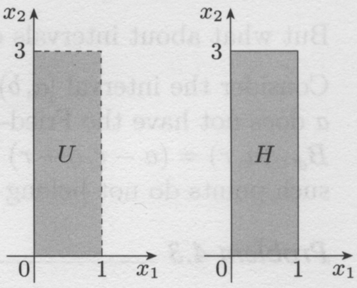


Figure 4.5

Solution

(a) For each $\mathbf{u} \in U$, we must find an $r > 0$ such that $B_{d^{(2)}}(\mathbf{u}, r) \subseteq U$.

Let $\mathbf{u} \in U$. Then

$$0 < u_1 < 1 \quad \text{and} \quad 0 < u_2 < 3,$$

and so we can find an $r > 0$ such that

$$0 \le u_1 - r < u_1 < u_1 + r \le 1 \quad \text{and} \quad 0 \le u_2 - r < u_2 < u_2 + r \le 3.$$

(We could take any r such that $0 < r \le \min\{u_1, 1 - u_1, u_2, 3 - u_2\}$.)

Now consider $B_{d^{(2)}}(\mathbf{u}, r)$. If $\mathbf{v} \in B_{d^{(2)}}(\mathbf{u}, r)$, then $d^{(2)}(\mathbf{u}, \mathbf{v}) < r$: that is,

$$\sqrt{(v_1 - u_1)^2 + (v_2 - u_2)^2} < r.$$

In particular, both

$$(v_1 - u_1)^2 < r^2 \quad \text{and} \quad (v_2 - u_2)^2 < r^2,$$

and so

$$|v_1 - u_1| < r \quad \text{and} \quad |v_2 - u_2| < r.$$

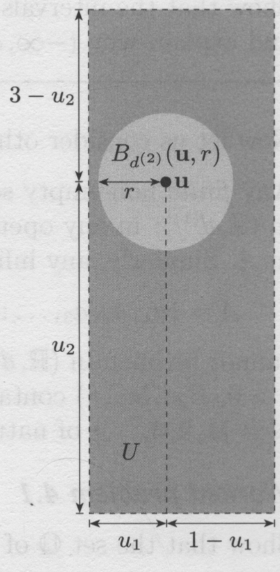


Figure 4.6

But this means that

$$0 \le u_1 - r < v_1 < u_1 + r \le 1 \quad \text{and} \quad 0 \le u_2 - r < v_2 < u_2 + r \le 3.$$

This shows that $\mathbf{v} \in U$, and hence $B_{d^{(2)}}(\mathbf{u}, r) \subseteq U$.

Since $\mathbf{u}$ is an arbitrary point of U, it follows that U is open.

(b) We need to find a point $\mathbf{h} \in H$ such that, no matter what $r > 0$ we choose, the ball $B_{d^{(2)}}(\mathbf{h}, r)$ is not contained within H. Since we already know that $(0,1) \times (0,3)$ is $d^{(2)}$-open, we investigate a point on the boundary of H. Consider the point $(1,2)$, for example, and suppose that $r > 0$. Then

$$d^{(2)}((1,2),(1+\tfrac{1}{2}r,2)) = \tfrac{1}{2}r < r,$$

and so $(1+\frac{1}{2}r, 2) \in B_{d^{(2)}}((1,2),r)$. But $(1+\frac{1}{2}r,2)$ is not in H, and so

$$B_{d^{(2)}}((1,2),r) \nsubseteq H.$$

Hence $(1,2)$ is a point in H that does not have the Fried-egg Property, and so H is not open. ■

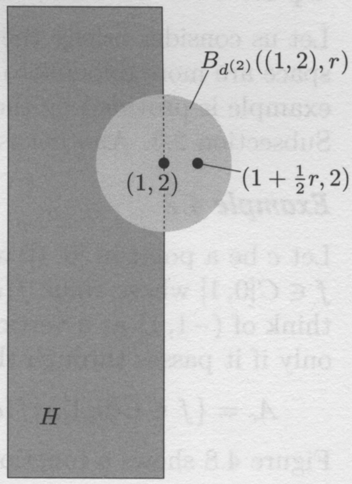


Figure 4.7

Remark

The results of Worked problem 4.2 can be generalized to show that any set of the form $(a,b) \times (c,d)$, often referred to as an *open rectangle*, is open in $(\mathbb{R}^2, d^{(2)})$ and that any set of the form $[a,b] \times [c,d]$, often referred to as a *closed rectangle*, is not open in $(\mathbb{R}^2, d^{(2)})$.

Problem 4.5

Consider the plane $\mathbb{R}^2$ with the product metric e_∞ obtained from the Euclidean metric $d^{(1)}$ on $\mathbb{R}$.

(a) Show that

$$U = (0,1) \times (0,3) = \{\mathbf{x} \in \mathbb{R}^2 : 0 < x_1 < 1, 0 < x_2 < 3\}$$

is an e_∞-open set.

(b) Show that

$$H = [0,1] \times [0,3] = \{\mathbf{x} \in \mathbb{R}^2 : 0 \le x_1 \le 1, 0 \le x_2 \le 3\}$$

is not an e_∞-open set.

Remark

The results of Problem 4.5 can be generalized to show that all open rectangles are open and all closed rectangles are not open in $\mathbb{R}^2$ with the product metric e_∞ obtained from the Euclidean metric $d^{(1)}$ on $\mathbb{R}$.

Open sets with the discrete metric

What are the open sets when the discrete metric is used? To answer this, we consider an arbitrary subset A of (X, d_0) and test its points for the Fried-egg Property; it turns out that *every subset of X is d_0-open.*

Problem 4.6

Let X be a set. Show that every subset of X is d_0-open.

Remark

We saw earlier that all non-empty finite sets $\{a_1, a_2, \ldots, a_n\}$, all infinite sets of the form $\{a_1, a_2, a_3, \ldots : a_1 < a_2 < a_3 < \cdots\}$ and the set $\mathbb{Q}$ of rational numbers are not open in the metric space $(\mathbb{R}, d^{(1)})$. However, it follows from the result of Problem 4.6 that each of these sets *is* open in the space $(\mathbb{R}, d_0)$. Thus a set can be open with respect to one metric but not with respect to another.

Open sets of continuous functions

Let us consider briefly the metric space $(C[0,1], d_{\max})$. Open sets in this space are more difficult to visualize than those in Euclidean spaces. One example is provided by the open balls in this space, discussed at the end of Subsection 2.3. Another is provided in the following example.

Example 4.2

Let c be a point in $[0,1]$, and consider the set A_c of all functions $f \in C[0,1]$ whose value $f(c)$ at c lies in the real open interval $(-1,1)$. If we think of $(-1,1)$ as a vertical 'gate' sited at $x = c$, then f belongs to A_c only if it passes through the gate:

$$A_c = \{f \in C[0,1] : |f(c)| < 1\}.$$

Figure 4.8 shows a function $g \in A_c$ and a function $h \notin A_c$.

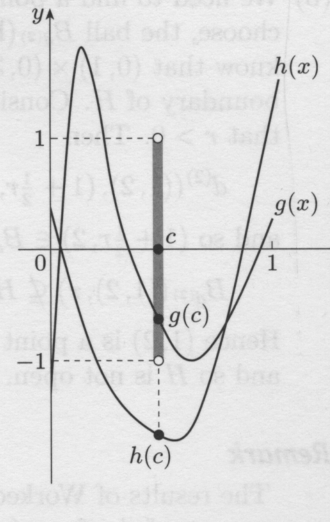


Figure 4.8

We now show that A_c is an open set in $(C[0,1], d_{\max})$.

Consider any $g \in A_c$. We shall show that g has the Fried-egg Property for $d_{\max}$ in A_c. Since $g(c) \in (-1,1)$, we can choose $r > 0$ with $r \leq 1 - |g(c)|$. Then,

$$(g(c) - r, g(c) + r) \subseteq (-1,1).$$

With this value of r, consider the open ball $B_{d_{\max}}(g,r)$. If $h \in C[0,1]$ belongs to this open ball, then $|h(x) - g(x)| < r$ for all $x \in [0,1]$, so that in particular $|h(c) - g(c)| < r$. Thus

$$-1 < g(c) - r < h(c) < g(c) + r < 1,$$

which shows that $h \in A_c$. Thus, if $h \in B_{d_{\max}}(g,r)$, then $h \in A_c$, proving that g has the Fried-egg Property for $d_{\max}$ in A_c.

Since g is an arbitrary point in A_c, it follows that A_c is an open subset of $(C[0,1], d_{\max})$. ■

It is easy to find sets that are not open in $(C[0,1], d_{\max})$. For example, consider the set S of all functions whose graphs pass through the origin:

$$S = \{f \in C[0,1] : f(0) = 0\}.$$

Figure 4.9 shows two functions f and g in S, and a function h not in S.

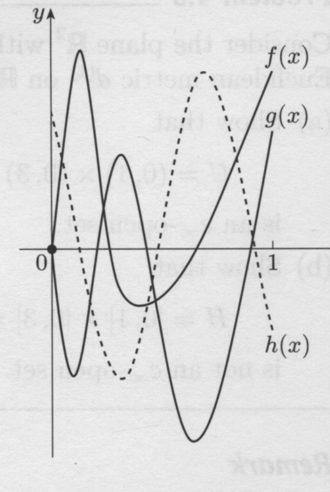


Figure 4.9

Problem 4.7

Show that the set S is not open.

Open sets and metrically equivalent metrics

Suppose that d_1 and d_2 are metrically equivalent metrics on a set X, so that there are positive real numbers m and M such that, for all $x, y \in X$,

$$md_1(x,y) \leq d_2(x,y) \leq Md_1(x,y).$$

Suppose further that U is d_1-open in X. Then U is also d_2-open. For, let $u \in U$. Since U is also d_1-open, there is an $r > 0$ such that $B_{d_1}(u,r) \subseteq U$. Put $s = mr$, and consider $B_{d_2}(u,s)$. Let $v \in B_{d_2}(u,s)$. Then

$$d_1(u,v) \leq \frac{1}{m} d_2(u,v) < \frac{s}{m} = \frac{mr}{m} = r.$$

So $v \in B_{d_1}(u,r) \subseteq U$. Since v is an arbitrary point of $B_{d_2}(u,s)$, we have $B_{d_2}(u,s) \subseteq U$ and hence U is d_2-open as predicted.

Similarly, if U is d_2-open then it is d_1-open.

We deduce that metrically equivalent metrics on a set X determine the *same* collection of open subsets of that set. In particular, since (by Theorem 3.3) the product metrics e_1, e_2 and e_∞ are metrically equivalent, they determine the *same* collection of open sets on the product space on which they are defined.

4.2 Continuity and open sets

In this subsection, we show that the continuity of a function between two metric spaces can be expressed entirely in terms of open sets. This is very important because, as we shall see in the next unit, it opens up the possibility of defining continuity in settings more general than metric spaces.

The crucial fact is that, if $f\colon X \to Y$ is (d,e)-continuous, then the inverse image of any e-open *ball* in Y is a d-open *subset* of X.

Lemma 4.2

Let (X,d) and (Y,e) be metric spaces and let $f\colon X \to Y$ be (d,e)-continuous on X. Then, for all $y \in Y$ and $r > 0$, $f^{-1}(B_e(y,r))$ is d-open.

Remark

This lemma tells us that the inverse image of any *open ball* under a continuous function is an *open set*, though not necessarily a ball.

Proof Suppose that $b \in f^{-1}(B_e(y,r))$. In order to show that $f^{-1}(B_e(y,r))$ is d-open, we must find a $\delta > 0$ such that

$$B_d(b,\delta) \subseteq f^{-1}(B_e(y,r)).$$

Since $b \in f^{-1}(B_e(y,r))$, we have $f(b) \in B_e(y,r)$. Therefore, by the Fried-egg Property for balls (Theorem 4.1), there is an open ball centred on $f(b)$ and contained entirely in $B_e(y,r)$. That is, there is an $r' > 0$ such that

$$B_e(f(b),r') \subseteq B_e(y,r).$$

Now let $c \in f^{-1}(B_e(f(b),r'))$. Then $f(c) \in B_e(f(b),r')$, so that $f(c) \in B_e(y,r)$; that is, $c \in f^{-1}(B_e(y,r))$.

Thus we have proved that

$$f^{-1}(B_e(f(b),r')) \subseteq f^{-1}(B_e(y,r)).$$

But f is continuous on X; in particular, f is continuous at b. Thus (putting $r' = \varepsilon$ in the inverse image definition of continuity), there is a $\delta > 0$ for which

$$B_d(b,\delta) \subseteq f^{-1}(B_e(f(b),r')) \subseteq f^{-1}(B_e(y,r)).$$

Since this is true of *any* $b \in f^{-1}(B_e(y,r))$, it follows that $f^{-1}(B_e(y,r))$ is d-open. ■

This lemma enables us to prove a theorem that expresses continuity entirely in terms of open sets.

Theorem 4.3

Let (X,d) and (Y,e) be metric spaces and suppose that $f\colon X \to Y$. Then f is (d,e)-continuous on X if and only if $f^{-1}(U)$ is a d-open subset of X whenever U is an e-open subset of Y.

Remark

This is a very interesting result: it shows that the continuity status of a function f between two metric spaces is determined entirely by whether the inverse image of an e-open set is d-open. In particular, we no longer have any mention of ε and δ.

Proof

Proof that f is (d,e)-continuous on X implies that $f^{-1}(U)$ is d-open whenever U is e-open

Suppose that f is (d,e)-continuous on X, and let U be an e-open subset of Y. We must show that $f^{-1}(U)$ is d-open.

Suppose that $a \in f^{-1}(U)$. To show that $f^{-1}(U)$ is d-open, we must find a $\delta > 0$ such that $B_d(a,\delta) \subseteq f^{-1}(U)$.

Since $a \in f^{-1}(U)$, it follows that $f(a) \in U$. Thus, since U is e-open, we can find an $r > 0$ such that

$$B_e(f(a),r) \subseteq U.$$

However, Lemma 4.2 implies that $f^{-1}(B_e(f(a),r))$ is d-open. Therefore, since $a \in f^{-1}(B_e(f(a),r))$, there is a $\delta > 0$ such that

$$B_d(a,\delta) \subseteq f^{-1}(B_e(f(a),r)).$$

But, since $B_e(f(a),r) \subseteq U$, we have

$$f^{-1}(B_e(f(a),r)) \subseteq f^{-1}(U),$$

and so

$$B_d(a,\delta) \subseteq f^{-1}(U).$$

Hence $f^{-1}(U)$ is d-open.

Proof that $f^{-1}(U)$ is d-open whenever U is e-open implies that f is (d,e)-continuous on X

Suppose that $f^{-1}(U)$ is d-open for all e-open sets U in Y. Let $a \in X$ and let $\varepsilon > 0$ be given. Then $B_e(f(a),\varepsilon)$ is an e-open subset of Y and so, by hypothesis, $f^{-1}(B_e(f(a),\varepsilon))$ is d-open. Now $a \in f^{-1}(B_e(f(a),\varepsilon))$ and so we can find a $\delta > 0$ such that

$$B_d(a,\delta) \subseteq f^{-1}(B_e(f(a),\varepsilon)).$$

Hence, by the inverse image definition of continuity, f is (d,e)-continuous at a.

Thus, since a is an arbitrary point of X, f is (d,e)-continuous on X. ■

Theorem 4.3 enables us to reformulate our definition of continuity for functions between metric spaces one more time, in terms of open sets.

Definition

Let (X,d) and (Y,e) be metric spaces.

A function $f\colon X \to Y$ is **continuous** on X if $f^{-1}(U)$ is a d-open subset of X whenever U is an e-open subset of Y.

Remarks

(i) Notice that this formulation defines continuity on the set X rather than at a point $a \in X$. Of course, if a function $f\colon X \to Y$ is continuous on X, then it is continuous at each point $a \in X$.

(ii) We shall refer to this definition of continuity on metric spaces as the **open set definition**.

Suppose two metrics d_1 and d_2 on a set X determine the same open sets: that is,

U is d_1-open if and only if U is d_2-open.

Then it does not matter which metric we use in testing for continuity: we always get the same result. We thus have the following result as a corollary to Theorem 4.3.

Corollary 4.4

Let X be a set, and let d_1 and d_2 be metrics on X.
Let (Y,e) be any metric space. If d_1 and d_2 determine the same open sets, then:

(a) if $f\colon X \to Y$ then

f is (d_1,e)-continuous on X if and only if f is (d_2,e)-continuous on X;

(b) if $g\colon Y \to X$ then

g is (e,d_1)-continuous on Y if and only if g is (e,d_2)-continuous on Y.

We have seen that metrically equivalent metrics determine the same open sets. Hence, for metrically equivalent metrics, Corollary 4.4 is equivalent to Theorem 3.4.

Remark

This corollary tells us that, if changing a metric on a set does not change the collection of open sets, then neither does it change the continuity status of any function.

Problem 4.8

Let (X, d) be a metric space and let d^* be the metric given by

$$d^*(x, y) = \min\{d(x, y), 1\}.$$

You may assume that d^* defines a metric.

(a) Show that, for $r > 0$,

$$B_{d^*}(x, r) = \begin{cases} B_d(x, r) & \text{if } 0 < r \leq 1, \\ X & \text{if } r > 1. \end{cases}$$

(b) Use the result in (a) and Corollary 4.4 to show that a function is (d, e)-continuous on X if and only if it is (d^*, e)-continuous on X.

4.3 Properties of open sets

We end this unit by investigating the general properties of open sets in metric spaces. Our starting point in *Unit A3* will be to use these properties to define a *topology*.

Our first result shows that the 'smallest' and 'largest' subsets of a metric space are always open.

Theorem 4.5

Let (X, d) be a metric space. Then both $\varnothing$ and X are d-open sets.

Proof First we consider the set $\varnothing$. Since there are no points in $\varnothing$, *every* point in $\varnothing$ has the Fried-egg Property for d in $\varnothing$. Hence, $\varnothing$ is d-open.

Now we consider the set X. Let $x \in X$ and let $r > 0$. Since $B_d(x, r) \subseteq X$, x has the Fried-egg Property for d in X. Hence X is d-open. ■

Now suppose that we are given two open subsets U and V of X. Must their intersection $U \cap V$ be open?

Theorem 4.6

In a metric space, the intersection of any two open sets is an open set.

Proof Let (X, d) be a metric space, and suppose that U and V are open subsets of X. Let $W = U \cap V$. We need to show that if $w \in W$, then it has the Fried-egg Property for d in W.

So suppose that $w \in W$; then $w \in U$ and $w \in V$. Since U and V are both open, there exist an $s > 0$ and a $t > 0$ such that

$$B_d(w, s) \subseteq U \quad \text{and} \quad B_d(w, t) \subseteq V.$$

Let $r = \min\{s, t\}$, so $r > 0$, and consider $B_d(w, r)$. If $x \in B_d(w, r)$, then

$$d(x, w) < r = \min\{s, t\},$$

and so $x \in B_d(w, s)$ and $x \in B_d(w, t)$. Thus

$$B_d(w, r) \subseteq B_d(w, s) \subseteq U \quad \text{and} \quad B_d(w, r) \subseteq B_d(w, t) \subseteq V.$$

This means that $B_d(w, r) \subseteq U \cap V = W$, and so w has the Fried-egg Property for d in W.

This proves that W is open. ■

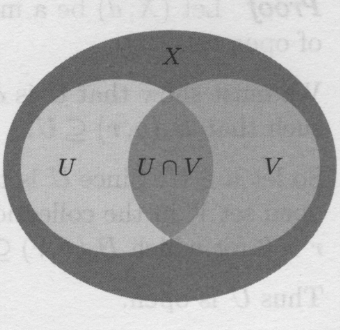


Figure 4.10

Remarks

(i) If U_1, U_2, U_3 are open, then

$$U_1 \cap U_2 \cap U_3 = (U_1 \cap U_2) \cap U_3$$

and we may use Theorem 4.6 first to deduce that $U_1 \cap U_2$ is open and then to deduce that $(U_1 \cap U_2) \cap U_3$ is open. Thus, the intersection of any three open sets is open.

This method extends to any *finite* intersection of open sets and allows us to conclude that, if $U_1, U_2, \ldots, U_n$ are open sets, then

$$U = U_1 \cap U_2 \cap \cdots \cap U_n$$

is also open.

(ii) However, this result does *not* extend to the intersection of *infinitely many* open sets. For example, consider the following collection of open intervals in $(\mathbb{R}, d^{(1)})$:

$$A_1 = (-1, 1), A_2 = (-\tfrac{1}{2}, \tfrac{1}{2}), A_3 = (-\tfrac{1}{3}, \tfrac{1}{3}), \ldots, A_n = (-\tfrac{1}{n}, \tfrac{1}{n}), \ldots.$$

Each set A_n (being an open interval) is open, but $A_1 \cap A_2 \cap A_3 \cap \cdots$ is the set of real numbers lying in *all* the intervals $(-\frac{1}{n}, \frac{1}{n})$, for all $n \in \mathbb{N}$, which is the single point $\{0\}$. But $\{0\}$ is not an open set, since any $d^{(1)}$-open ball about 0 contains points other than 0.

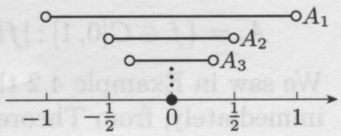


Figure 4.11

Open sets have another property that turns out to be remarkably useful. Figure 4.12 shows the union of two open balls in the Euclidean plane. This union is *not* an open ball, but (as illustrated) it does have the Fried-egg Property, so it is an open set. This is generally true: the union of open balls is an open set. Moreover, things do not stop there: the union of *any* collection of open sets is an open set, irrespective of whether the collection is finite or infinite.

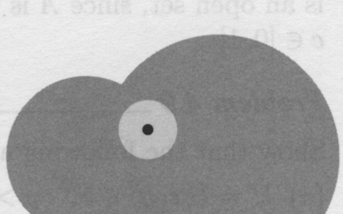


Figure 4.12

> ***Theorem 4.7***
>
> In a metric space, the union of any collection of open sets is open.

Proof Let (X,d) be a metric space, and let U be the union of a collection of open sets.

We must show that U is d-open: that is, for each $u \in U$, there is an $r > 0$ such that $B_d(u,r) \subseteq U$.

So let $u \in U$. Since U is the union of a collection of open sets, there is an open set V in the collection such that $u \in V$. Since V is d-open, there is an $r > 0$ for which $B_d(u,r) \subseteq V$. But $V \subseteq U$ and hence $B_d(u,r) \subseteq U$.

Thus U is open. ■

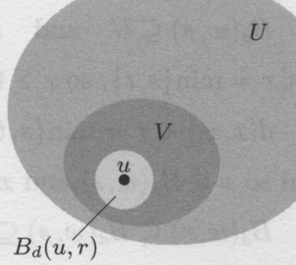


Figure 4.13

One consequence of Theorem 4.7 is that, if we can write a given set as a union of open sets, then we immediately know that it is open and we do not have to find open balls around each point to verify this. For example, let

$$J = \{(x,0) \in \mathbb{R}^2 : -2 \leq x \leq 2\},$$

and let

$$U = \{\mathbf{u} \in \mathbb{R}^2 : d^{(2)}(\mathbf{u},\mathbf{j}) < 1 \text{ for some } \mathbf{j} \in J\}.$$

Thus U is a rectangle with rounded ends, and it may be intuitively clear to you that U is open. We could construct a proof of this from the definition of an open set, but there is no need: because U has been defined as the union of the infinite collection of open balls $B_{d^{(2)}}((x,0),1)$ where $-2 \leq x \leq 2$, we can deduce immediately from Theorem 4.7 that U is open.

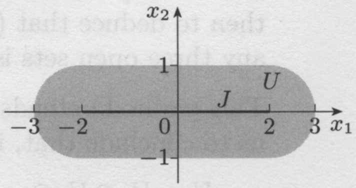


Figure 4.14

This principle is also effective for more abstract metric spaces, such as $(C[0,1], d_{\max})$. For example, let $c \in [0,1]$ and consider the following subset of $(C[0,1], d_{\max})$:

$$A_c = \{f \in C[0,1] : |f(c)| < 1\}.$$

We saw in Example 4.2 that this is an open set. We can deduce immediately, from Theorem 4.7, that the set

$$A = \{f \in C[0,1] : \text{there is a } c \in [0,1] \text{ for which } |f(c)| < 1\}$$

is an open set, since A is the union of the infinite collection of sets A_c for $c \in [0,1]$.

Problem 4.9

Show that the following are open sets in the metric space $(\mathbb{R}^2, d^{(2)})$.

(a) $U = \{(x,y) \in \mathbb{R}^2 : y > 1\}$

(b) $V = \{(x,y) \in \mathbb{R}^2 : (x-1)^2 + (y-1)^2 < 1 \text{ and } (x-2)^2 + (y-2)^2 < 1\}$

(c) The set W of points $(x,y) \in \mathbb{R}^2$ such that

$$(x-a)^2 + (y-b)^2 < \tfrac{1}{4}$$

for some (a,b) with $a^2 + b^2 = 1$.

Solutions to problems

1.1 (a) $e_1((0,0),(1,0)) = |1-0| + |0-0| = 1$;
(b) $e_1((0,0),(0,1)) = |0-0| + |1-0| = 1$;
(c) $e_1((0,1),(1,0)) = |1-0| + |0-1| = 2$.

1.2 It follows from (M1) that

$$B_d(a,0) = \{x : d(a,x) < 0\} = \varnothing,$$

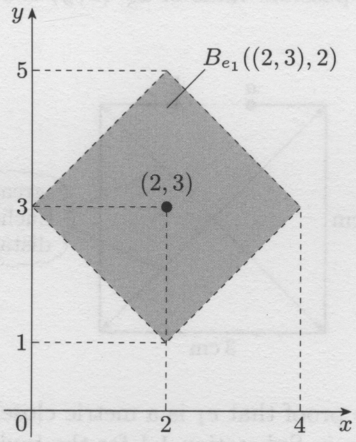
$$B_d[a,0] = \{x : d(a,x) \le 0\} = \{x : d(a,x) = 0\} = \{a\},$$

$$S_d(a,0) = \{x : d(a,x) = 0\} = \{a\}.$$

1.3 The problem is to find all $\mathbf{x}$ satisfying

$$e_1((2,3),\mathbf{x}) = |x_1 - 2| + |x_2 - 3| < 2.$$

The point $(2,3)$ is the centre of the ball, and its radius is 2. If we imagine the origin of coordinates at $(2,3)$ and put the vertices of the diamond 2 units from the centre along the axes, the required figure results.

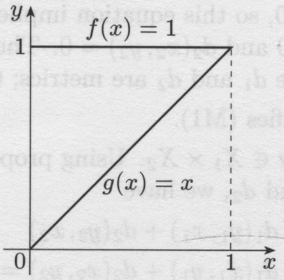


2.1 The function d is not a metric. It does not satisfy (M3), the Triangle Inequality, since

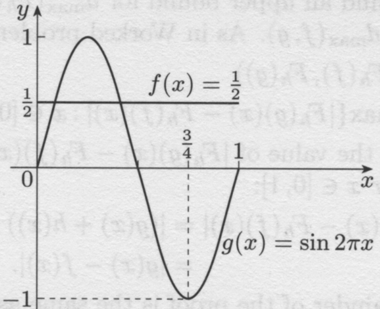
$$d(x,y) + d(y,z) = 1 + 2 = 3 < 4 = d(x,z).$$

2.2 $d_C(\mathbf{a},\mathbf{e}) = 2^{-2} = \frac{1}{4}$, $d_C(\mathbf{b},\mathbf{e}) = 2^{-3} = \frac{1}{8}$, $d_C(\mathbf{c},\mathbf{e}) = 2^{-4} = \frac{1}{16}$, $d_C(\mathbf{d},\mathbf{e}) = 2^{-5} = \frac{1}{32}$.

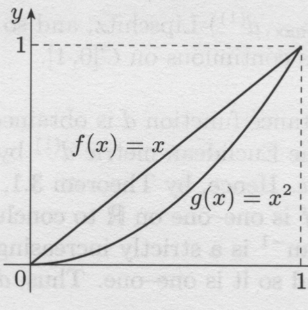
2.3 (a) Here the maximum value of $|g(x) - f(x)|$ is 1, when $x = 0$; so $d_{\max}(f,g) = 1$.

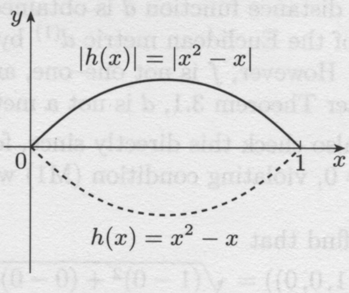


(b) In this case, the furthest $g(x)$ gets from $\frac{1}{2}$ is when $g(x) = -1$ (at $x = \frac{3}{4}$) and so $d_{\max}(f,g) = |(-1) - \frac{1}{2}| = \frac{3}{2}$.



(c) On $[0,1]$, $g(x) \le f(x)$ with equality at the endpoints only. Thus the maximum of $|g - f|$ over $[0,1]$ occurs at the minimum of the function $h(x) = g(x) - f(x) = x^2 - x$ over this domain. This occurs at the point x_0 for which the derivative $h'(x) = 2x - 1$ vanishes, so that $x_0 = \frac{1}{2}$. Thus $d_{\max}(f,g) = |\frac{1}{4} - \frac{1}{2}| = \frac{1}{4}$.





2.4 $d_{\min}$ does not define a metric on $C[0,1]$: it does not satisfy (M1). For example, if $f(x) = 0$ and $g(x) = x$ for $x \in [0,1]$, then $d_{\min}(f,g) = 0$ but $f \ne g$.

2.5 (a) $F_c(f)(x) = x^2 + 1$.

(b) $F_c(f)(x) = \sin x + \pi$. (Note that this is $(\sin x) + \pi$ and *not* $\sin(x + \pi)$.)

2.6 The proof is very similar to that of Worked problem 2.2.

Let $h \in C[0,1]$. Let $f \in C[0,1]$ and consider the continuity of F_h at f.

Let $\varepsilon > 0$ be given. We must find a $\delta > 0$ such that, for all $g \in C[0,1]$,

$$d_{\max}(F_h(f), F_h(g)) < \varepsilon \quad \text{whenever} \quad d_{\max}(f,g) < \delta.$$

We first find an upper bound for $d_{\max}(F_h(f), F_h(g))$ in terms of $d_{\max}(f,g)$. As in Worked problem 2.2,

$$\begin{aligned} &d_{\max}(F_h(f), F_h(g)) \\ &\quad = \max\{|F_h(g)(x) - F_h(f)(x)| : x \in [0,1]\}. \end{aligned}$$

Consider the value of $|F_h(g)(x) - F_h(f)(x)|$ for a *particular* $x \in [0,1]$:

$$\begin{aligned} |F_h(g)(x) - F_h(f)(x)| &= |(g(x) + h(x)) - (f(x) + h(x))| \\ &= |g(x) - f(x)|. \end{aligned}$$

The remainder of the proof is the same as in Worked problem 2.2 except that F_h replaces F_c and $h \in C[0,1]$ replaces $c \in \mathbb{R}$.

2.7 We show that F is $(d_{\max}, d^{(1)})$-Lipschitz, and deduce that it is continuous.

Let $f, g \in C[0,1]$. We need to find an upper bound for $d^{(1)}(F(f), F(g))$ in terms of $d_{\max}(f,g)$. Now

$$\begin{aligned} d^{(1)}(F(f), F(g)) &= |g(0) - f(0)| \\ &\le \max\{|g(x) - f(x)| : x \in [0,1]\} \\ &= d_{\max}(f,g). \end{aligned}$$

Hence F is $(d_{\max}, d^{(1)})$-Lipschitz, and so by Theorem 2.5 is continuous on $C[0,1]$.

3.1 The distance function d is obtained from the pull-back of the Euclidean metric $d^{(1)}$ by the function $f(x) = \tan^{-1} x$. Hence, by Theorem 3.1, it is enough to show that f is one–one on $\mathbb{R}$ to conclude that d is a metric. But $\tan^{-1}$ is a strictly increasing function (see Figure 3.2) and so it is one–one. Thus, d is a metric.

3.2 The distance function d is obtained from the pull-back of the Euclidean metric $d^{(1)}$ by the function $f(x) = x^2$. However, f is not one–one, and so, by the remark after Theorem 3.1, d is not a metric.

(You can also check this directly since, for all $x \in \mathbb{R}$, $d(x, -x) = 0$, violating condition (M1) when $x \neq 0$.)

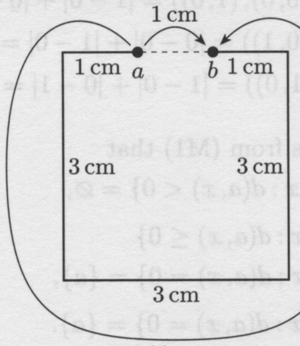
3.3 We find that

$$d^{(3)}(\mathbf{0}, (1,0,0)) = \sqrt{(1-0)^2 + (0-0)^2 + (0-0)^2} = 1,$$
$$d^{(3)}(\mathbf{0}, (0,0,1)) = \sqrt{(0-0)^2 + (0-0)^2 + (1-0)^2} = 1,$$

and so $(1,0,0)$ and $(0,0,1)$ are in S.

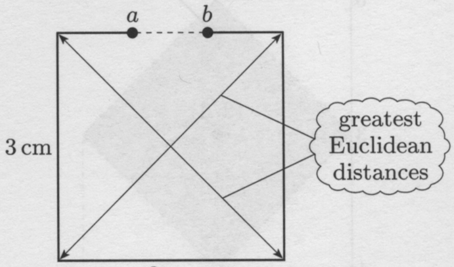
$$\begin{aligned} &d_S^{(3)}((1,0,0),(0,0,1)) \\ &\quad = \sqrt{(0-1)^2 + (0-0)^2 + (1-0)^2} = \sqrt{2}. \end{aligned}$$

3.4 **(a)** The string is 11 cm long and a and b are at its ends, so $d(a,b) = 11$.



The perimeter of a square of side 3 cm is 12 cm, so the gap between the ends has length 1 cm. Therefore $d_S^{(2)}(a,b) = 1$.

(b) For the Euclidean metric, the points at opposite corners of the square are furthest apart. Their distance apart is $\sqrt{3^2+3^2} = 3\sqrt{2}$ cm. So the maximum possible value of $d_S^{(2)}(x,y)$ is $3\sqrt{2}$.



3.5 The proof that e_1 is a metric closely follows the proof given in Subsection 1.1 for the taxicab metric.

(M1) For all $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$, $d_1(x_1, y_1)$ and $d_2(x_2, y_2)$ are non-negative, since d_1 and d_2 are metrics, and hence so is their sum: thus $e_1(\mathbf{x}, \mathbf{y}) \ge 0$.

For all $\mathbf{x} \in X_1 \times X_2$,

$$e_1(\mathbf{x}, \mathbf{x}) = d_1(x_1, x_1) + d_2(x_2, x_2) = 0 + 0 = 0,$$

since d_1 and d_2 are metrics.

Conversely, suppose $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$ are such that $e_1(\mathbf{x}, \mathbf{y}) = 0$. Then

$$0 = d_1(x_1, y_1) + d_2(x_2, y_2).$$

Since d_1 and d_2 are metrics, $d_1(x_1, y_1) \ge 0$ and $d_2(x_2, x_2) \ge 0$, so this equation implies that $d_1(x_1, y_1) = 0$ and $d_2(x_2, y_2) = 0$. Thus $x_1 = y_1$ and $x_2 = y_2$, since d_1 and d_2 are metrics; that is, $\mathbf{x} = \mathbf{y}$.

Thus e_1 satisfies (M1).

(M2) Let $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$. Using property (M2) for the metrics d_1 and d_2, we have

$$\begin{aligned} e_1(\mathbf{y}, \mathbf{x}) &= d_1(y_1, x_1) + d_2(y_2, x_2) \\ &= d_1(x_1, y_1) + d_2(x_2, y_2) = e_1(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Thus e_1 satisfies (M2).

(M3) Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X_1 \times X_2$. Then, using property (M3) for the metrics d_1 and d_2,

$$\begin{aligned} e_1(\mathbf{x}, \mathbf{z}) &= d_1(x_1, z_1) + d_2(x_2, z_2) \\ &\leq (d_1(x_1, y_1) + d_1(y_1, z_1)) \\ &\quad + (d_2(x_2, y_2) + d_2(y_2, z_2)) \\ &= (d_1(x_1, y_1) + d_2(x_2, y_2)) \\ &\quad + (d_1(y_1, z_1) + d_2(y_2, z_2)) \\ &= e_1(\mathbf{x}, \mathbf{y}) + e_1(\mathbf{y}, \mathbf{z}). \end{aligned}$$

Thus e_1 satisfies (M3).

Therefore e_1 is a metric on $X_1 \times X_2$.

3.6
$$\begin{aligned} e_1(\mathbf{0}, (2,1)) &= |2-0| + |1-0| = 3; \\ e_2(\mathbf{0}, (2,1)) &= \sqrt{(2-0)^2 + (1-0)^2} = \sqrt{5}; \\ e_\infty(\mathbf{0}, (2,1)) &= \max\{|2-0|, |1-0|\} = 2. \end{aligned}$$

Since these distances are all different, we conclude that the three metrics are distinct.

3.7 **(a)** For the unit open ball for e_∞ centred at the origin, we have to find those $(x_1, x_2) \in \mathbb{R}^2$ satisfying

$$\max\{|x_1|, |x_2|\} < 1.$$

This is satisfied provided $-1 < x_1 < 1$ and $-1 < x_2 < 1$, and the points satisfying these inequalities form the interior of the square shown in Figure 3.5.

(b) The unit closed balls centred at the origin are the regions shown in Figure 3.5 together with their boundary points, comprising the bounding diamond, circle and square.

The unit spheres centred at the origin are simply the bounding diamond, circle and square of the regions shown in Figure 3.5.

3.8 Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$. Then

$$\begin{aligned} e(\mathbf{x}, \mathbf{y}) &= |y_1 - x_1| + 2|y_2 - x_2| \\ &\leq 2|y_1 - x_1| + 2|y_2 - x_2| \\ &= 2(|y_1 - x_1| + |y_2 - x_2|) = 2e_1(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Similarly,

$$\begin{aligned} e_1(\mathbf{x}, \mathbf{y}) &= |y_1 - x_1| + |y_2 - x_2| \\ &\leq |y_1 - x_1| + 2|y_2 - x_2| \\ &= e(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Thus,

$$e_1(\mathbf{x}, \mathbf{y}) \leq e(\mathbf{x}, \mathbf{y}) \leq 2e_1(\mathbf{x}, \mathbf{y}).$$

Hence e and e_1 are metrically equivalent.

3.9 **(a)** (i) $(0, (0,0,0,0,0,0,\ldots))$.
(ii) $(0, (1,0,1,0,1,0,\ldots))$.
(Notice the extra brackets.)

(b) Consider the composites of the projection functions with f, that is, $p_1 \circ f : \mathbf{C} \to \mathbb{R}$ and $p_2 \circ f : \mathbf{C} \to \mathbf{C}$. Now

$$(p_1 \circ f)(\mathbf{x}) = x_1 = g(\mathbf{x}),$$

where $g : \mathbf{C} \to \mathbb{R}$ is given by $g(\mathbf{x}) = x_1$. By assumption, g is $(d_{\mathbf{C}}, d^{(1)})$-continuous. Also,

$$(p_2 \circ f)(\mathbf{x}) = (x_2, x_3, x_4, \ldots) = \sigma(\mathbf{x}),$$

and we saw in Worked problem 2.1 that σ is $(d_{\mathbf{C}}, d_{\mathbf{C}})$-continuous.

Since both $p_1 \circ f$ and $p_2 \circ f$ are continuous, it follows from Theorem 3.7 that f is $(d_{\mathbf{C}}, e)$-continuous, where e is any of e_1, e_2 or e_∞.

4.1 **(a)** $f^{-1}(\{0\}) = \{x \in \mathbb{R} : x^2 = 0\} = \{0\}$.

(b) $f^{-1}([0,4]) = \{x \in \mathbb{R} : x^2 \in [0,4]\} = [-2,2]$.

(c) $f^{-1}([1,4)) = \{x \in \mathbb{R} : x^2 \in [1,4)\}$
$= (-2,-1] \cup [1,2)$.

(d) $f^{-1}((-4,-1)) = \{x \in \mathbb{R} : x^2 \in (-4,-1)\} = \varnothing$.

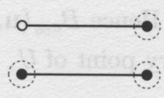
4.2 Let $a, x \in X$ with $d(a,x) > r$. Let $s = d(a,x) - r > 0$. Let $y \in X$ satisfy $d(x,y) < s$. Then, using the Reverse Triangle Inequality (Theorem 1.1) and property (M2), we have

$$\begin{aligned} d(a,y) &\geq |d(x,a) - d(x,y)| \\ &= |d(a,x) - d(x,y)| \\ &= d(a,x) - d(x,y) \\ &> d(a,x) - s = r. \end{aligned}$$

That is, $d(a,y) > r$. Thus, if $y \in B_d(x,s)$, then $y \notin B_d(a,r)$, and hence

$$B_d(a,r) \cap B_d(x,s) = \varnothing.$$

4.3 These two sets fail to be open due to the presence in each set of at least one endpoint. In $(a,b]$, the point b does not have the Fried-egg Property. In $[a,b]$, neither a nor b has the Fried-egg Property.



$(a,b]$

$[a,b]$

4.4 To show that (a,∞) is open, let $x \in (a,\infty)$. Then $x > a$, so $x - a > 0$. Take $r < x - a$; then $B_{d^{(1)}}(x,r) \subseteq (a,\infty)$. So each point $x \in (a,\infty)$ has the Fried-egg Property and hence (a,∞) is open.

Similarly, let $y \in (-\infty,a)$. Then $y < a$, so $a - y > 0$. Take $s < a - y$; then $B_{d^{(1)}}(y,s) \subseteq (-\infty,y)$, and so $(-\infty,a)$ is open.

Next, let $z \in \mathbb{R}$. For any $r > 0$, $B_{d^{(1)}}(z,r) \subseteq \mathbb{R}$, and so $(-\infty,\infty)$ is open.

The intervals $(-\infty,a]$ and $[a,\infty)$ are not open because of the presence in each set of the endpoint a, which does not have the Fried-egg Property.

4.5 (a) For each $\mathbf{u} \in U$, we must find an $r > 0$ such that $B_{e_\infty}(\mathbf{u}, r) \subseteq U$.

Let $\mathbf{u} \in U$. Then

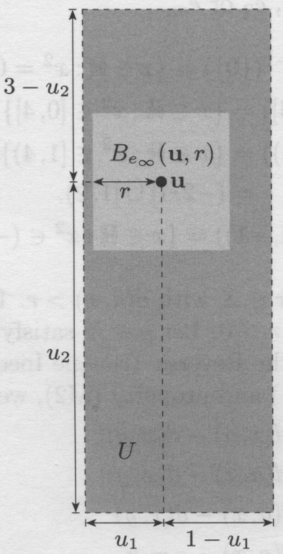
$$0 < u_1 < 1 \quad \text{and} \quad 0 < u_2 < 3,$$

and so we can find an $r > 0$ such that

$$0 \leq u_1 - r < u_1 < u_1 + r \leq 1,$$
$$0 \leq u_2 - r < u_2 < u_2 + r \leq 3.$$

(We could take any r such that

$$0 < r \leq \min\{u_1, 1 - u_1, u_2, 3 - u_2\}.)$$



Now consider $B_{e_\infty}(\mathbf{u}, r)$. If $\mathbf{v} \in B_{e_\infty}(\mathbf{u}, r)$, then $e_\infty(\mathbf{u}, \mathbf{v}) < r$: that is,

$$\max\{|v_1 - u_1|, |v_2 - u_2|\} < r.$$

In particular, both

$$|v_1 - u_1| < r \quad \text{and} \quad |v_2 - u_2| < r.$$

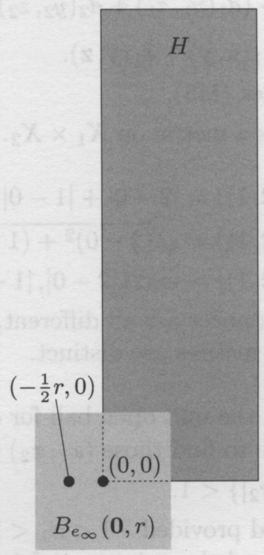
But this means that

$$0 \leq u_1 - r < v_1 < u_1 + r \leq 1,$$
$$0 \leq u_2 - r < v_2 < u_2 + r \leq 3,$$

showing that $\mathbf{v} \in U$. Hence $B_{e_\infty}(\mathbf{u}, r) \subseteq U$.

Since $\mathbf{u}$ is an arbitrary point of U, we conclude that U is open.

(b) We need to find a point $\mathbf{h} \in H$ such that, no matter what $r > 0$ we choose, the ball $B_{e_\infty}(\mathbf{h}, r)$ is not contained within H. Since we already know that $(0,1) \times (0,3)$ is e_∞-open, we investigate a point on the boundary of H. Consider the point $\mathbf{0} = (0,0)$, for example, and suppose that $r > 0$.



Observe that

$$e_\infty((0,0), (-\tfrac{1}{2}r, 0)) = \tfrac{1}{2}r < r$$

and so $(-\frac{1}{2}r, 0) \in B_{e_\infty}(\mathbf{0}, r)$. But $(-\frac{1}{2}r, 0)$ is not in H and so

$$B_{e_\infty}(\mathbf{0}, r) \nsubseteq H.$$

Hence $\mathbf{0}$ is a point in H that does not have the Fried-egg Property, and so H is not open.

4.6 Let $A \subseteq X$ be any subset and let $a \in A$. We must show that a has the Fried-egg Property for d_0 in A. To do this, we have to find an $r > 0$ such that $B_{d_0}(a, r) \subseteq A$.

Choose any $0 < r \leq 1$. We know from Subsection 2.1 that, for such an r,

$$B_{d_0}(a, r) = \{a\} \subseteq A.$$

So a has the Fried-egg Property for d_0. Hence, since a is an arbitrary point of A, A is a d_0-open set.

4.7 Let $f_0 \in C[0,1]$ be the function defined by

$$f_0(x) = 0 \quad \text{for } x \in [0,1].$$

Then $f_0 \in S$. We show that f_0 does not have the Fried-egg Property for $d_{\max}$ in S.

Consider *any* $r > 0$. Let $g \in C[0,1]$ be the function defined by

$$g(x) = \tfrac{1}{2}r \quad \text{for } x \in [0,1].$$

Then $d_{\max}(f_0, g) = \frac{1}{2}r < r$, and so $g \in B_{d_{\max}}(f_0, r)$. But $g \notin S$, and so f_0 does not have the Fried-egg Property for $d_{\max}$ in S. Thus, S is not open.

4.8 **(a)** First suppose that $r > 1$. Since $d^*(x, y) \le 1$ for all $y \in X$, it follows that

$$B_{d^*}(x, r) = X.$$

Now suppose that $0 < r \le 1$. If $y \in B_d(x, r)$ then $d(x, y) < r \le 1$, and so $d^*(x, y) = d(x, y) < r$. Thus $y \in B_{d^*}(x, r)$ and so

$$B_d(x, r) \subseteq B_{d^*}(x, r).$$

Similarly, if $y \in B_{d^*}(x, r)$ then $d^*(x, y) < r \le 1$, and so $d(x, y) = d^*(x, y) < r$. Thus $y \in B_d(x, r)$ and so

$$B_{d^*}(x, r) \subseteq B_d(x, r).$$

Hence, for $0 < r \le 1$,

$$B_{d^*}(x, r) = B_d(x, r).$$

(b) If U is d-open then, for each $u \in U$, we can find an $r > 0$ such that $B_d(u, r) \subseteq U$. Take $s = \min\{r, 1\}$. Then, using the result in (a),

$$B_{d^*}(u, s) = B_d(u, s) \subseteq B_d(u, r) \subseteq U.$$

So U is d^*-open.

Conversely, if U is d^*-open then, for each $u \in U$, we can find an $r > 0$ such that $B_{d^*}(u, r) \subseteq U$. From (a), $B_{d^*}(x, r) = B_d(u, r)$ or X, so that certainly $B_d(u, r) \subseteq B_{d^*}(u, r)$, and hence

$$B_d(u, r) \subseteq U.$$

So U is d-open.

Thus the d-open and d^*-open subsets of X are the same and so, by Corollary 4.4, a function is (d, e)-continuous on X if and only if it is (d^*, e)-continuous on X.

4.9 **(a)** For each $(x, y) \in U$, $y - 1 > 0$. Put $r_{x,y} = y - 1$ and consider the open balls $B_{d^{(2)}}((x, y), r_{x,y})$. U is the union of these infinitely many open balls, for all $(x, y) \in U$, and so is the union of an infinite collection of open sets. Therefore, by Theorem 4.7, U is open.

(b) $V = B_{d^{(2)}}((1, 1), 1) \cap B_{d^{(2)}}((2, 2), 1)$; that is, V is the intersection of two unit open balls. Therefore, by Theorem 4.6, V is open.

(c) The point (x, y) belongs to W if and only if it is at a distance less than $\frac{1}{2}$ from some point (a, b) on the unit circle. That is, W is the union of an infinite collection of open balls of radius $\frac{1}{2}$. Therefore, by Theorem 4.7, W is open.

Index